A DEEP-LEARNING-BASED REAL-TIME OBJECT DETECTION TECHNIQUE IN RAILWAY TRANSPORTATION

BY

Md Sojib Molla
ID: 201-15-3729

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Ms. Nasima Islam Bithi
Lecturer
Department of CSE
Daffodil International University

Co-Supervised By

Nusrat Khan
Lecturer
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

26 JANUARY 2024

APPROVAL

This Project titled "**A deep-learning-based real-time object detection technique in railway transportation**", submitted by Md Sojib Molla to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 26/01/2023

BOARD OF EXAMINERS

Dr. Md. Ismail Jabiullah — **Chairman**
Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Raja Tariqul Hasan Tusher — **Internal Examiner 1**
Assistant Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Md. Abbas Ali Khan — **Internal Examiner 2**
Assistant Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Dr. Md. Zulfiker Mahmud — **External Examiner 1**
Associate Professor, External Member
Department of CSE
Jagannath University

DECLARATION

I therefore, declare that, this project has been done by me under the supervision of **Ms. Nasima Islam Bithi, Lecturer, Department of CSE** Daffodil International University. I also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

for

Ms. Nasima Islam Bithi
Lecturer
Department of CSE
Daffodil International University

Co-Supervised by:

for

Nusrat Khan
Lecturer
Department of CSE
Daffodil International University

Submitted by:

Sojib

Md Sojib Molla
ID: 201-15-3729
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, I express my heartiest thanks and gratefulness to almighty God for His divine blessing makes me possible to complete the final year project successfully.

I am really grateful and wish me profound my indebtedness to **Ms. Nasima Islam Bithi**, **Lecturer**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of my supervisor in the field of "Deep Learning" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stage have made it possible to complete this project.

I would like to express my heartiest gratitude to Professor **Dr. Sheak Rashed Haider Noori**, Head**,** Department of CSE, for his kind help to finish my project and also to other faculty member and the staff of CSE department of Daffodil International University.

I would like to thank my entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, I must acknowledge with due respect the constant support and patients of my parents.

ABSTRACT

In computer vision, object detection is the process of finding or identifying things in pictures or videos. It provides a quick and effective method for classifying objects. Its extensive applications in object recognition, separation, and detection have advanced safety and convenience in our contemporary life.

In this paper, we introduce a model that detects the object in real time at railway crossings. In the experiments, I used YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x of the YOLOv5 family. Each model was trained separately to see which one performed better in terms of recall, mAP@0.5, mAP@0.5:0.95, and precision. To train and validate the object detection system, I used open-source images captured on railway crossings and also used online data. The dataset consists of 949 images containing eight classes, almost 550 are raw data which is captured on railway crossings and almost 400 images are collected online. The dataset I used divided includes 758 images for training and 189 images for validation. After the experiment of YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5x model, YOLOv5x obtained the highest average of precision at 89.4%, recall at 85.5%, and mAP at 95.6%. So YOLOv5x is the most stable method followed by YOLOv5s, YOLOv5m, and YOLOv5l. The experimental results that YOLOv5x algorithms are accurately detecting real-time objects in railway crossings. It will be a great help in reducing the collisions of trains and also reducing the accident rate by real-time object detection.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 Introduction

Railway crossings are vital intersections in the world's transportation networks that link railroad tracks with roads or pathways for pedestrians. Because of their importance in allowing automobile and rail traffic to mix peacefully, it is crucial to make sure these crossroads are safe. Modern technologies, especially object detection models, must be integrated to improve operational effectiveness, safety, and monitoring at railroad crossings. Computer vision techniques enable object identification at railway crossings by analyzing different items present in the crossing area. For control systems and decision-making procedures, real-time vehicle, pedestrian, and potential barrier detection and classification provide crucial data. The application of the YOLOv5 (You Only Look One Level) item identification model in this technological setting is the particular emphasis of this work. YOLOv5, which is well-known for its accuracy and efficiency, provides an efficient method of processing images, guaranteeing quick detection without sacrificing accuracy. Crossing monitoring has always been dependent on rule-based systems or manual surveillance, both of which had difficulty quickly adjusting to changing conditions and novel scenarios. A noteworthy development is the addition of the YOLOv5 model, which leverages deep learning capabilities to quickly and independently identify objects in the crossing area. By improving item detection accuracy and facilitating railway operators' ability to make decisions in real-time, this methodology enhances overall safety and efficiency in railway systems. Our study's main goal is to investigate the nuances of object recognition at train crossings using the cutting-edge YOLOv5 model. The main objective of this massive project is to advance railway safety technologies to unprecedented levels, with a focus on addressing the particular difficulties brought about by these crucial junctions where rail and road networks converge. We recognize that the successful deployment of an advanced object detection system has the potential to completely transform the concept of railway safety and open the door to a transportation network that complies with contemporary safety and reliability standards. As such, our focus goes beyond simple technological innovation. Our goal is to change the way that railway safety is seen and handled at important traffic intersections such as railroad crossings by utilizing the YOLOv5 model. Because of the convergence of several means of transportation, these crossroads are focus locations where potential risks develop, making them particularly difficult to manage safely. The YOLOv5 model, which is well-known for its precision and effectiveness, is a useful instrument in our efforts to improve object recognition accuracy at these crucial points, which will ultimately improve overall safety. A successful advanced object identification system has far-reaching effects that go well beyond the boundaries of technology. Visioning a time where efficiency and safety coexist harmoniously, this kind of technology might coordinate a total overhaul of railroad crossing operations, ushering in a new era of increased operational efficiency and safety standards. Our study aims to be a key player in bringing about this revolutionary change. Utilizing state-of-the-art object detection technology to lower accident risks is the fundamental goal at the center of this change. Our research aims to reshape the field of safety procedures by expeditiously and accurately identifying different entities that cross railroads, such as cars, pedestrians, and potential impediments. In addition to risk reduction, a global standard for railroad crossing safety is being emphasized, which is paving the way for a paradigm change in the transportation industry. Our research takes a broad approach that considers a variety of safety factors, such as vehicle traffic, pedestrian movements, and potential obstacles. Our goal is to build a solid foundation that puts dependability and safety first throughout the transportation system by integrating advanced object detection technology. This foundation aims to surpass current safety regulations and create new standards by surpassing current expectations and demonstrating real-time detecting skills and precision. Essentially, the goal of our research is to catalyze a comprehensive reinterpretation of safety regulations within the transportation industry. The proposed global standard goes beyond conventional safety protocols by integrating technological advancements that not only reduce hazards but also improve the safety environment at train crossings proactively. This comprehensive strategy not only foresees possible obstacles but also puts the transportation system in a position to adjust and react quickly to new situations, guaranteeing a degree of reliability and safety that smoothly fits in with the constantly changing requirements of modern safety standards. Our ultimate goal is to help build a transportation system that genuinely emphasizes the safety and dependability needed for a sustainable and secure future, while also embracing technological improvements.

1.2 Motivation

The impetus for this study project is the need to address safety issues at railroad crossings, which are important locations where railroad tracks converge with roadways or pedestrian walkways. To avoid mishaps and maintain the smooth operation of rail and road traffic, these intersections require close attention. The impetus for this study stems from a profound understanding of the intrinsic difficulties presented by conventional monitoring techniques in guaranteeing the security and functionality of these crucial transportation hubs. Recognizing the importance of these issues strengthens our resolve to investigate cutting-edge technical solutions that have the potential to enhance safety precautions and change the dynamics of railroad crossings. The limits of using traditional methods emphasize the need for improved object-detecting technologies in modern railway operations. If we consider the last 5 years, we can see many train accidents due to the mismanagement of the rail crossing. 8 major accidents can be seen.

Figure 1.2.1: Train collision in Bangladesh

The use of rule-based systems or manual surveillance has shown to be inadequate in addressing the changing issues that dynamic settings at railroad crossings present. The complexities of contemporary railway systems, where a more responsive and intelligent approach is required due to the interplay of multiple factors, are beyond the scope of these traditional methodologies. The shortcomings of the current monitoring systems become increasingly obvious as train networks grow and experience greater traffic. The increasing need for effective and automated safety measures at railroad crossings has made real-time, on-demand object detection services even more urgent. The inadequacies in the existing object detection approaches are not limited to technological impediments; they also represent concrete problems that affect the overall safety and operational effectiveness of railroad crossings. Advanced systems are strategically necessary to meet the increasingly complicated demands of today's railway infrastructure, rather than being a purely technological novelty. Taking this into account, our work aims to fill more than just a technical vacuum. It seeks to be a driving force behind breakthroughs in the industry by actively participating in the creation of cutting-edge object detection technologies that are painstakingly crafted to address the complex problems unique to railroad crossings. We want to close this gap by using creative solutions, ushering in a new era of efficiency and safety in railway operations, and creating a model for intelligent and adaptable systems within the larger framework of transportation infrastructure.

I have purposefully selected the YOLOV5 (You Only Look One Level) model as the cornerstone of our study because of its proven speed and accuracy in processing images while preserving good detection performance. To support decision-making and control systems, railway crossings require the quick recognition and classification of a variety of things, including cars, pedestrians, and possible impediments. Our purposeful selection of the YOLOV5 model to anticipate the revolution in railway crossing management and surveillance represents a strategic commitment to utilizing state-of-the-art technology. As a key component in our efforts to improve railway operations safety, the YOLOV5 model stands out for its ability to offer real-time insights into the dynamic environment of railroad crossings. The deliberate choice of this model represents more than just a technical choice; it represents a proactive strategy meant to herald in a new era of accuracy and efficiency in object identification and classification at these crucial transportation crossroads. This research is significant because it goes beyond the immediate objective of improving safety; it is a ground-breaking endeavor with broad implications. The development of a dependable and efficient object detection system for railroad crossings has the potential to drastically lower accident rates and promote improved operational procedures. Such a system has a "ripple effect" that affects not only the immediate safety improvement but also the general safety and efficacy of railway systems more broadly. A more secure and effective transportation network is ensured by our research, which is focused on providing train operators with real-time information, a crucial component of railway safety. Moreover, the broader implications of this research go beyond the particular field of train safety. This project aims to be a pathfinder in the disciplines of computer vision and intelligent systems, rather than just tackling urgent problems. Through this research, we may better comprehend object detection in dynamic, complex contexts such as railway crossings. In addition to advancing object detection technology, the investigation of novel neural network topologies, improvement of data augmentation methods, and testing of transfer learning approaches in the context of railway crossing object detection hold potential for use in related fields. These findings, which originate from the basic research of object detection in railroad crossings, could affect several domains, such as traffic management, autonomous vehicles, and critical infrastructure surveillance. Thus, the project aims to be a catalyst for larger advancements in the domains of computer vision and intelligent systems, rather than just a means of resolving a specific problem. To put it briefly, the goal of this research is to increase the safety of railroad crossings by employing cutting-edge object detection technologies, particularly the YOLOV5 model. By developing a tailored system, the goal is to usher in a new era of efficiency and accuracy in the identification and classification of items at railway crossings. In addition to providing quick fixes for the safety laws of trains, this research has the potential to significantly impact related fields and progress object detection in general. The ultimate objective is to build a transportation network that is technologically advanced, safer, and more dependable while smoothly conforming to contemporary standards for dependability and safety at train crossings.

1.3 Rationale of the study

The primary driving force behind this study is the deliberate and exclusive use of the YOLOV5 model for the crucial assignment of object detection at railroad crossings. Dedicated to supporting safety protocols in railroad operations, this research deliberately focuses its dependence on the unique features of the YOLOv5 model. By purposefully focusing just on one specific model, the study seeks to maximize its strengths and develop an extremely accurate and effective object recognition system that is specially designed for the complex and dynamic environment of railroad crossings. The rationale behind choosing the YOLOv5 model as

- The central component of this study is its established ability and efficacy to process photos quickly without sacrificing detection performance standards.
- The adaptability of the model becomes especially relevant in the real-time requirements of railway crossings, where it is critical to quickly and accurately identify a variety of elements, such as cars, people, and potential impediments.
- Advance safety protocols within railway environments and contribute to the ongoing evolution of object detection technology in the transportation sector.
- In terms of fast photo processing without sacrificing a high level of detection performance.
- Various objects including vehicles and potential obstacles are instantly and accurately identified.
- The model's ability to recognize a variety of things quickly and precisely fits very well with the intricate operational needs of railroad crossings, where a moment's decision-making can have a big impact on both efficiency and safety.
- The research's overarching goal is to not only address the immediate challenges presented by railway crossings but also to initiate a transformative shift in the monitoring and management of these intersections, ushering in a new era of enhanced safety and operational efficiency.

- It is a calculated choice to capitalize on and enhance its unique advantages in the object identification domain.
- The model's remarkable proficiency in maneuvering through intricate sceneries, its flexibility in responding to varying lighting conditions, and its ability to distinguish between various object sizes are well-suited to the difficult obstacles presented by railroad crossings.
- This exclusive focus is a purposeful strategy that aims to improve safety at railroad crossings while also carrying a deep-seated desire to be a key contributor to the development and improvement of the YOLOv5 model.
- Through the full utilization of the model's remarkable capabilities, the research endeavors to create a highly efficient and precise system that skillfully addresses the particular difficulties presented by railway environments.
- To sum up, the purpose of this study is to investigate why the YOLOv5 model is intentionally used for object recognition at train crossings.
- Through utilizing all of this model's amazing potential, the research aims to create a highly precise and effective system that can handle the unique difficulties that come with working in train contexts.
- The deliberate choice to focus the research on the YOLOv5 model is an obvious example of a methodical approach that was carefully thought out to make use of and enhance the model's special advantages in object detection.
- By fully using the model's potential, the study aims to develop a system that not only satisfies but exceeds the requirements of railway environments, offering increased precision and effectiveness.
- The study's purposeful use of the YOLOv5 model is a clear indication of its steadfast commitment to actively leading a shift in railway safety paradigms rather than just solving present problems.

1.4 Research question

1. Which earlier research has been done to address the difficulties with object detection at railroad crossings, with an emphasis on the approaches, tools, and results of these studies?

2. What tactics and methods have been utilized in earlier studies to guarantee the precision and dependability of object detection systems at railroad crossings, with a focus on the methods for confirming the accuracy of items that have been identified?

3. Is there any raw picture data available that is pertinent to the study of object detection at railroad crossings? If yes, how big is this dataset regarding the quantity and variety of things photographed in a railroad crossing setting?

4. What specific categories are included in the purview of this detection-based initiative for railroad crossings, and how are these categories established for the detection of objects at these vital intersections for transportation?

5. What distinguishes the selected characteristic or aspect of the research's suggested technique from previous approaches in the area of object identification in railway crossing scenarios, making it special or unique?

6. What effects do the suggested roles and techniques in this study have on bettering overall transportation network operations, and how do they improve railway safety and efficiency, especially in supporting real-world scenarios?

1.5 Expected Outcome

Developing and validating an advanced object recognition system at railway crossings using the cutting-edge YOLOv5 model is the main goal of this project, which aims to achieve previously unheard-of levels of accuracy and efficiency. The main objective is to dramatically improve safety at railroad crossings by implementing state-of-the-art technology that skillfully handles the dynamic problems present in these vital traffic locations. The expected result is the development of an efficient YOLOv5-based object identification model specifically designed for the complex and dynamic environment of railroad crossings. The anticipated outcomes encompass the proficient recognition and classification of diverse entities existing at railroad crossings, including but not restricted to automobiles, pedestrians, and possible hindrances. It is anticipated that the YOLOV5 model will outperform conventional techniques, offering a noticeably higher degree of object recognition accuracy. Furthermore, the system is expected to have strong generalization skills, correctly recognizing objects in a variety of contexts with various lighting and weather patterns. The YOLOv5 model's real-time capabilities play a critical role in enabling a smooth integration into the current railway infrastructure, guaranteeing rapid interventions and prompt responses to improve overall safety. Anticipated is the creation of a user-friendly interface that can be accessed on several platforms, such as computers and mobile devices. It is anticipated that this technology will provide real-time insights into the dynamic environment of railroad crossings, helping to effectively control traffic and reduce the risk of accidents. In addition, the project intends to produce extensive datasets with annotations related to item detection in railroad crossings. These datasets will provide important insights to the larger research community working on improving railway safety and optimizing transportation infrastructure, in addition to acting as benchmarks for assessing the model's effectiveness. To sum up, the project's expected results go beyond just successfully deploying the YOLOv5-based object detection system at railroad crossings. To set new benchmarks for railway safety technology and significantly develop transportation infrastructure, the emphasis is on attaining increased safety, real-time efficiency, and seamless integration into varied platforms.

1.6 Project Management and Finance

Using the YOLOv5 model for object detection at railway crossings, the research was started by gathering photographs from several websites. Images were carefully chosen to depict the variety of scenarios and difficulties related to these important transportation intersections, given the unique focus on railway crossings. The YOLOv5 model was selected because it is well-known for its effectiveness in real-time object identification and was thought to be appropriate for handling the challenges of recognizing objects at railway crossings, including cars, pedestrians, and possible obstructions.

The YOLOv5 model was trained on a dataset of photos sourced from various web sources, which ensured a thorough portrayal of actual railway crossing scenarios. The research was able to access a variety of circumstances and scenarios by utilizing publicly available photographs from the internet, which improved the training data and increased the adaptability of the model. In terms of project funding, there were no direct expenses incurred in the use of internet photos or in training the YOLOv5 model. By using publicly available photographs, costs were kept to a minimum and the team could concentrate on improving and developing the object detection algorithm. The cost-effectiveness of using pre-trained models and cloud-based services for model training was also advantageous to the project. This strategy emphasizes a thoughtful and effective use of the resources at hand, with an emphasis on reducing cost effects while meeting the goals of the object detection project for railroad crossings. The choice to utilize the YOLOv5 model and obtain photographs from the internet is indicative of a practical project management approach that seeks to provide dependable outcomes with minimal financial outlay.

Table 1.6.1: Project Management Table

Work	Time
Data Collection Review	1 month
Papers and Articles	3 months
Experimental Setup	1 month
Implementation	1 month
Report Writing	2 months
Total	8 months

1.7 Report Layout

This research report's format was thoughtfully created to methodically convey the extensive discoveries and understandings gained from the study of object detection at railway crossings using the YOLOv5 model, using pictures from many online sources. As an introduction, Chapter 1 gives a summary of the research topic and highlights the importance of using the YOLOv5 model to improve object recognition at railroad crossings. Chapter 2, findings a methodology of related work. The study methodologies utilized are covered in detail in Chapter 3, which provides an overview of the data collection tactics and statistical tools that were employed. Chapter 4 summarizes the main findings of the study, which include the YOLOv5 model's performance and the empirical findings obtained from the different photographs obtained from the internet.

In Chapter 5, the research's possible effects on transportation management and railway safety are examined, along with how the results might affect the intersections' larger context. The analysis and conclusion, which summarizes the major discoveries and their applications, are finally presented in Chapter 6. This chapter provides a thorough summary of the research findings, analyzes the benefits and drawbacks of the YOLOv5 model in this particular application, and suggests future directions for the field of object recognition for railway crossings using online photo data.

CHAPTER 2

BACKGROUND

2.1 Preliminaries/Terminologies

In the domain of object detection, there has been a substantial leap forward, particularly with the integration of advanced models like YOLOv5. The amalgamation of machine learning and deep learning techniques has proven pivotal in elevating the capabilities of object detection systems. Recent studies and projects, particularly those focused on enhancing safety at railway crossings, have delved into the application of YOLOv5, a state-of-the-art object detection model renowned for its exceptional speed and accuracy. These initiatives entail the collection of images from diverse internet sources to construct a comprehensive dataset, crucial for the training of the YOLOv5 model. Object detection takes center stage in these endeavors, playing a crucial role in identifying and localizing various elements within the complex environment of railway crossings to ensure safety and prevent potential accidents. YOLOv5's distinctive features, including real-time detection and precise bounding box predictions, position it as a valuable asset in overcoming the challenges associated with railway safety. As we embark on an exploration of this research, a comprehensive understanding of these preliminary concepts lays a solid foundation for grasping the intricacies and advancements inherent in applying YOLOv5 to object detection within the context of railway crossings.

2.2 Related Works

Kasper-Eulaers et al. (2021) investigated the use of YOLOv5 to identify large trucks at rest stops in the winter, enabling real-time parking space occupancy prediction. Google Colaboratory (Colab) was used to train the model; it offers free, no-configuration access to powerful GPUs. Roboflow.ai created a notebook with a pre-trained COCO weight and YOLOv5 as its foundation. Before overfitting after roughly 150 epochs, the model rapidly improved in terms of precision, recall, and mean average precision. The validation data's box, objectness, and classification losses similarly rapidly declined until approximately epoch 15. The trained algorithm can detect the front cabin of heavy-duty cars with high confidence, according to the results; nevertheless, it appears to be more difficult to detect the rear, especially when it is placed far from the camera. [18]

Malta et al. (2021) presented a deep learning neural network-based task assistant model. Some of the components of a car were recognized using the YOLOv5 network. Eight different types of parts were identified in the dataset that was created, which included 582 images from three videos with similar lighting conditions. These parts included the oil dipstick, battery, engine oil reservoir, wiper water tank, air filter, brake fluid reservoir, coolant reservoir, and power steering reservoir. Each frame's photos were transformed to a 416 × 416 format, which is required as input by the selected architecture. During the development process, mobile phones were utilized for taking images and movies, while PCs were utilized to execute software. A laptop computer with access to a Google Colab virtual machine was used to train the object identification model. For the two models (YOLOv5s and YOLOv5m), the accuracy was comparable to what other authors had found for issues of a similar kind. In the test sets, YOLOv5s showed that they could reliably identify eight distinct mechanical elements in a vehicle engine with high precision, and they consistently had recall rates over 96.8%. These findings are nearly identical to those of the bigger model. Findings show that the network is competent and quick enough to be used to support the identification of an automobile's component parts. [19]

A self-attention mechanism-based YOLOv5 model for polyp target identification was proposed by Wan et al. in 2021. The mosaic method was used in the data preprocessing stage to enhance the amount of training data in the data set; cross-stage partial networks (CSPNet) were used as the backbone network to extract the information features in the image, which solved the problem of gradient disappearance; and the feature pyramid architecture with attention mechanisms was used to enhance the detection performance of varying-size polyps. Stochastic gradient descent (SGD) and backpropagation were used to train the suggested approach end-to-end using a cloud computing infrastructure outfitted with eight 16-GB GPUs, a 16-core CPU, and 64 GB of RAM. YOLOv5 employed a new feature pyramid network (FPN) structure that improved the bottom-up path, which improved the propagation of low-level features; a path aggregation network (PANET) as the neck for feature aggregation; and spatial pyramid pooling (SPP) to improve the model's detection of objects with different scales. The author's approach produced outstanding results. The accuracy, recall rate, and F1score in the Kvasir-SEG data set were 0.915, 0.899, and 0.907, respectively. The accuracy, recall, and F1 score in the WCY dataset were 0.913, 0.921, and 0.917, respectively. In particular, this technique significantly decreased the false positive rate by employing full-image data to estimate the target window utilizing each network. [20]

A defect identification model based on YOLOv5 was created by Yao et al. (2021) and has the ability to quickly and reliably identify faults. To help the model identify minor flaws, a little object detection layer was included. To improve the accuracy of the regression, the squeeze-and-excitation (SE) layer and the loss function complete intersection over union (CIoU) were included. The cosine annealing technique was utilized to enhance the model's performance after it was trained via transfer learning. Comparing the mAP@0.5 of YOLOv5 to the original method, the improvement was about 9%, reaching 94.7%. [21]

An end-to-end real-time helmet identification technique for motorcycle riders based on the YOLOv5 algorithm was presented by Jia et al. in 2021. The K-means method in the COCO dataset determines the initial anchor box size of YOLOv5, and the suggested HFUT-MH dataset differs significantly from the COCO dataset. This approach attained a mAP of 97.7%, an F1-score of 92.7%, and frames per second of 63, which surpassed previous state-of-the-art detection methods created by Yao et al. (2021) and has the ability to quickly and reliably identify faults. To help the model identify minor flaws, a little object detection layer was included. To improve the accuracy of the regression, the squeeze-and-excitation (SE) layer and the loss function complete intersection over union (CIoU) were included. The cosine annealing technique was utilized to enhance the model's performance after it was trained via transfer learning. Comparing the mAP@0.5 of YOLOv5 to the original method, the improvement was about 9%, reaching 94.7%. [22]

Liu et al. (2021) presented a real-time method for detecting railway signal lights using YOLOv5. To demonstrate the efficacy of the suggested approach, experiments were carried out. A dataset of signal-light-equipped subway scenes was created, and the YOLOv5 model was trained on it. Running at 100 frames per second, the YOLOv5s-trained signal light detection model has an accuracy and recall rate of 0.972 on average.

Deep learning methods, in particular, have been employed and evaluated more and more in recent years (Liu et al., 2019) for fruit detection. Because machine learning is more precise and dependable than older methods, it offers a better answer to problems like occlusion and green tomato detection. The problem of green tomato detection is rarely studied because of how hard it is to segment green tomatoes and tell them apart from the backdrop due to their similar color. [27]

Lü et al. (2014) used a support vector machine (SVM) that was trained only on the RGB color space to recognize branches and fruit in natural environments. They reported that their approach achieved an accuracy of 92.4% for fruits, outperforming previous threshold-based strategies. The outcomes were, nevertheless, subject to lighting influences. [28]

The algorithm for counting fruits was applied by Rahnemoonfar & Sheppard (2017) and is based on Resnet and Inception. In real-time, the algorithm's overall prediction performance was 91%. Even if some of the fruits overlap under the shadow, the computer might still be able to count them. Nevertheless, the technique was not employed for detection; it was just used to count fruits. [29]

Córdova et al. conducted numerous experiments to examine the performance of YOLOv5, RetinaNet, EfficientDet, Faster R-CNN, and Mask R-CNN on litter identification based on the TACO dataset [23].

In related work, rubbish categorization and detection models were constructed in rural areas using the YOLOv5s architecture [24]. The suggested model included an attention combination technique to create a more robust background network topology. To create a more complete representation of the suggested model, the process outlined in [24] combined channel and spatial dimensions as feature information. This method made it possible to include a wider variety of data and specifics, which could improve the model's efficacy and accuracy in a number of object-detecting applications. The results showed that the mAP at 0.5, accuracy, and recall were 96.4%, 91.1%, and 93.5%, respectively.

Lin et al. conducted a prior work aimed to enhance the architecture of YOLOv5 by adding a feature map attention (FMA) at the backbone layer's conclusion [25]. This change was a major step forward for the field because it made it possible to include more information and specifics, which might improve the model's efficacy and accuracy across a range of object detection applications. The suggested model's capacity to extract features and identify eight waste categories of floating objects was enhanced by FMA. To generate real-scene environment settings, the model combined the tagged target objects with the background pictures of the clear river. The model achieved a mAP@0.5 of 79.41% on the testing dataset, as per the findings.

Zang et al. conducted a supervised attention mechanism that was added to YOLOv5m in research to enhance its feature extraction capability [26]. To enhance trash recognition from photos and videos in an actual situation, the model was coupled with a multimodal knowledge network. The model's mAP@0.5 of 72.8% was attained as a result of the combination, surpassing the performance of 72.8% of the original YOLOv5m.

2.3 Comparative Analysis and Summary

A few studies on various object detection have previously been conducted. The comparison of those research works with their algorithm and result analysis are below:

Table 2.3.1: Summary of related works.

SI	Authors Name	Algorithm	Findings
1	Kasper-Eulaers et al., 2021 [18]	Real-time detection of heavy-duty vehicles for occupancy of parking spots using YOLOv5	Improvement in Precision, recall, and mAP in detection. The model was able to detect the front cabin with high confidence but failed when located at a far distance.
2	Malta et al., 2021 [19]	Recognition of different constituents of a car using the YOLOv5 series.	The model was able to achieve precision and recall of 96.8% in the detection of the eight parts of the car engine used in the experiment compared with larger models used for the same purpose

3	Wan et al., 2021 [20]	Experiment using the YOLOv5 model for a self-attention mechanism for polyp target detection.	The model was able to achieve excellent recall, precision, and accuracy of above 90% due to the use of full image information during prediction in each network.
4	Yao et al., 2021 [21]	The experiment of defect detection using YOLOv5 with an addition of a Squeeze-and-Excitation (SE) layer.	The model achieved a mAP@0.5 of 94.7% which improved the model by 9% compared to the original model
5	Jia et al., 2021 [22]	A real-time end-to-end helmet detection of motorcyclists using YOLOv5 using K-means algorithm to calculate the anchors	The model achieved 97.7% mAP and 92.7% F1 scores which outperforms other state-of-the-art models
6	Cordova et al., 2022 [23]	YOLOv5	YOLOv5 outperformed other CNN architecture in terms of speed and accuracy in detecting litter detection applications
7	Jiang et al., 2022 [24]	YOLOv5s	YOLOv5s-based garbage classification and detection model in rural areas improved background network structure by adding an attention-combining mechanism, leading to better feature representation and potentially boosting the accuracy and efficiency of the model. High recall, mAP@0.5, and

			accuracy rates were attained using the model.
8	Lin et al., 2021 [25]	YOLOv5s	The model achieved a mAP@0.5 of 79.41% on the testing dataset
9	Zang et al., 2021 [26]	YOLOv5m	YOLOv5m improved feature extraction ability and improved garbage detection in real-scene images and videos, achieving a mAP@0.5 of 72.8%
10	Rahnemoonfar & Sheppard, 2021 [29]	Resnet and Inception	The prediction performance of 91% in real-time and also the algorithm is able to count the fruits.

2.4 Scope of problem

The goal of this research project is to create a reliable object detection system that is specially designed for railroad crossings by utilizing the state-of-the-art computer vision architecture known as the YOLOv5 model. The main goal is to use YOLOv5's high precision and real-time capabilities to quickly and reliably detect a variety of items in the dynamic and complicated environment of railroad crossings. Because of its well-known effectiveness in object recognition, the YOLOv5 model is a great choice for handling the unique problems that come with railroad crossings. These difficulties include a variety of objects kinds, including cars, people, and animals; changing lighting, and different viewpoints recorded by security cameras and other monitoring equipment. The project's scope includes the careful modification and optimization of the YOLOv5 model to improve its performance in the context of railroad crossings. This entails fine-tuning the model to accurately identify and classify items, guaranteeing that it functions flawlessly in real-time scenarios that are critical to railway safety and operational effectiveness. The thorough training of the YOLOv5 model on a variety of datasets that replicate the unpredictability of railway crossing scenarios is a crucial component of the research. The system seeks to achieve robust generalization by subjecting the model to a range of situations, object kinds, and environmental factors. This will enable the model to function well in a variety of railway crossing locations. One of the primary factors in determining the project's goals is practical application. The goal is to implement the YOLOv5-based object detection system on various monitoring devices that are frequently used in railway infrastructure, such as unmanned aerial vehicles (UAVs) and surveillance cameras. The designed system's flawless integration into the current railway surveillance networks is ensured by this deployment. One of the primary factors in determining the project's goals is practical application. The goal is to implement the YOLOv5-based object detection system on various monitoring devices that are frequently used in railway infrastructure, such as unmanned aerial vehicles (UAVs) and surveillance cameras. The designed system's flawless integration into the current railway surveillance networks is ensured by this deployment. It is important to understand that the project does not address traffic management or railway operations in general; rather, its concentration is on the implementation of the YOLOv5 model for object detection. The project aims to improve safety and operational effectiveness in these crucial transportation areas by addressing the complex difficulties related to object detection in railway crossings by utilizing the special characteristics of YOLOv5.

2.5 Challenges

Researchers are currently tackling the complex problems related to object detection in this crucial domain in response to the urgent demand for enhanced accessibility and safety in railroad crossings. The wide range of things that are present, including cars, pedestrians, and other obstructions, creates a challenging environment for creating reliable and broadly useful object detection models. The changing lighting conditions, varied perspectives, and variety in item appearances acquired by surveillance cameras need the use of advanced techniques like computer vision and machine learning. To build complete object recognition systems for railroad crossings, researchers must master the challenging challenge of smoothly integrating numerous modalities, which include object forms, movements, and the impact of environmental conditions. To successfully combine these many elements, sophisticated techniques must be used. This will result in systems that are accurate and dependable and can recognize things in real-time situations. A major challenge in this effort is the lack of diversified and high-quality object detection datasets tailored to railway crossings. One major obstacle to achieving the best possible performance of models in different settings is the lack of complete and well-labeled data. To achieve real-time object detection, systems must be able to quickly and correctly identify objects while taking into consideration changes and contextual factors that could affect the relevance of items that are identified. The process becomes even more complex because of unpredictable movements, background noise, and potential obstacles in the railway crossing environment. In light of these difficulties, it is critical to guarantee the robustness and flexibility of object detection models since outside factors may interfere with the detection procedure. Researchers are unwavering in their dedication to developing object-detecting technology for railroad crossings despite these difficulties. Their main goal is to consistently identify and classify objects in real-time circumstances, thereby making a significant contribution to transportation safety. The desire to build a transportation system that is safer and more inclusive is what motivates this commitment. Overcoming the obstacles in object detecting technology could lead to improved railway operations efficiency in addition to accident prevention. In the end, researchers see a world where everyone can travel more easily and safely, signifying a noticeable advancement in the state of railroad safety as a whole.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Research Subject and Instrumentation

Identifying objects at railroad crossings is the main goal of my research project. The unique difficulties that come with item classification in this setting stem from the variety of railway environments, which add layers of complexity that require careful consideration. To classify items in this dynamic environment, it becomes necessary to have a large dataset that includes a variety of events that can occur at railroad crossings. The goal of this project is to create an advanced deep-learning model that is especially suited for object detection at railroad crossings. This study's practical application made use of the Python programming language and well-known libraries, such as NumPy, Scikit-learn (Sklearn), and OpenCV. The Windows operating system was selected as the research environment because it offered an appropriate platform for analysis and experimentation. Every step of the training and testing process took place in a Jupyter notebook, which made experimentation more efficient and allowed for a thorough analysis of the final model. The difficulties that come with object detection at railroad crossings include the variety of object appearances, variances in lighting, and the possibility of occlusions or background interference. Because of these intricacies, deep learning techniques must be used to reliably identify and categorize items in real-time settings, guaranteeing the effectiveness and safety of railway operations. Through a direct approach to these problems and the application of deep learning techniques, this study aims to build a strong object detection model that can reliably identify a wide range of items seen at railroad crossings. A comprehensive grasp of the model's accuracy and dependability is ensured by a careful assessment of its performance using pertinent evaluation measures. Furthermore, the study conducts comprehensive testing and validation to examine the model's efficacy in real-world circumstances. The research hopes to make a substantial contribution to the field of object detection at railroad crossings with this all-encompassing method, which will ultimately improve safety protocols and operational effectiveness in railroad transit. It is anticipated that the created model will be essential in reducing hazards and promoting a safer, more welcoming train environment. The object detection model that has been developed is expected to have a major influence on reducing risks and promoting a more secure and hospitable training environment. Through efficient item recognition and classification at railway crossings, the model has the potential to significantly lower potential dangers and accidents.

3.2 Data Collection Procedure

To guarantee the diversity and richness of the dataset, a thorough data-gathering procedure was carefully crafted for the object detection model in railway crossings. Images were taken from a variety of channels to accurately depict the nuances of actual train surroundings. At railroad crossings, surveillance cameras were placed strategically to offer high-resolution photos that captured dynamic circumstances in real-time. To add a variety of situations to the dataset, databases about railroad crossings and transportation safety that are accessible to the public were examined. This enhanced the object identification model's resilience by modeling various illumination situations, meteorological scenarios, and object orientations. A thorough preprocessing process was applied to every image, which included scaling to a standard 512x512 pixel dimension. An important step was the annotation phase, which involved carefully naming objects like cars, CNG, and any impediments to make sure the model training procedure that followed was informed. To ensure that there were no discrepancies, quality control procedures were strictly followed. A portion of the annotated photos were manually inspected to confirm that the bounding box annotations and object labels were accurate. The final dataset consists of a heterogeneous set of photos that represent different kinds of weather, different lighting situations, and different object appearances seen at railroad crossings. The purpose of this intentional heterogeneity is to provide strong generalization capabilities to the object identification model so that it can recognize things in the diverse and ever-changing railway surroundings.

To train and validate the object detection system, I used open-source images captured on railway crossings and also used online data. Out of 949 images, almost 550 are raw data which is captured on railway crossings and almost 400 images are collected from online. The dataset I used divided includes 758 images for training and 189 images for validation.

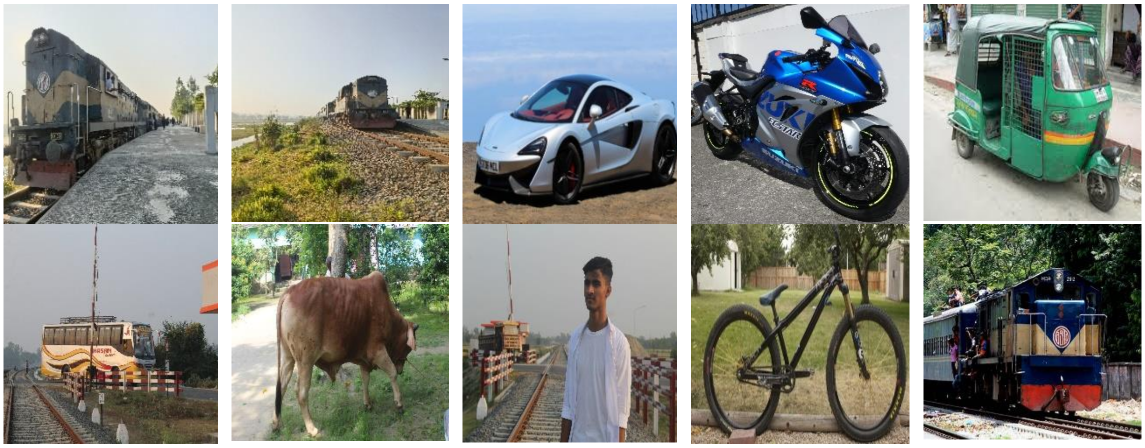


Figure 3.2.1: Sample images from the dataset

Figure 3.4.3 provides a visual representation of the dataset, the size, and the distribution of the object. (i) displays the object detection frame after normalizing the image size. (ii) displays the ratio of the object detection frame to the image size.

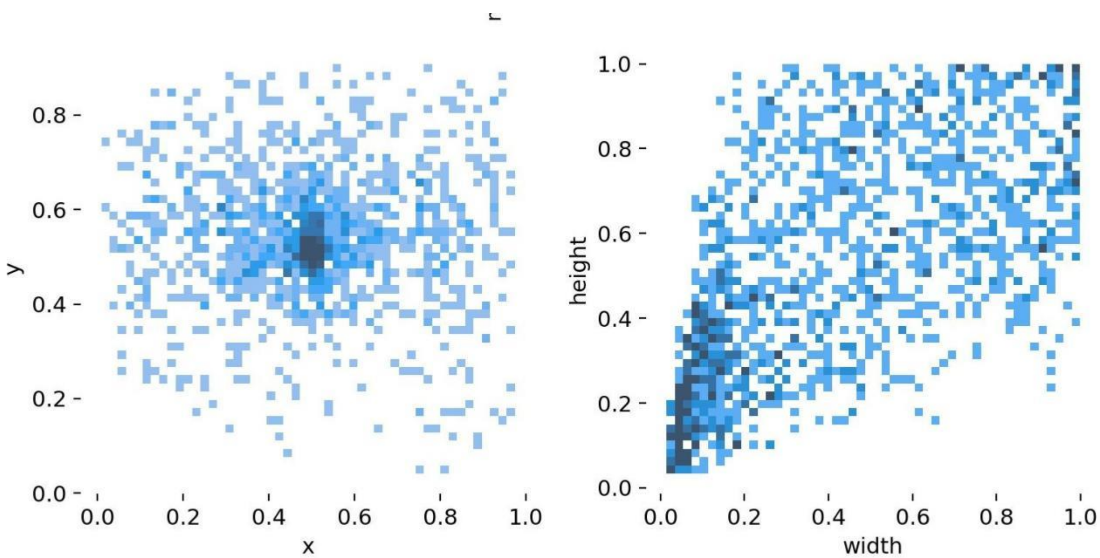


Figure 3.2.2: Data visualization (i) The location of object boxes (ii) The size of object boxes

3.3 Statistical Analysis

This research set out to create a comprehensive and nuanced model to improve object detection skills specifically for railway crossings. The dataset used for this study included nine different classes that were carefully selected to capture the wide variety of items and situations seen in the complex terrain of railroad systems. These groups, which reflected the complexity of railway crossings, comprised automobiles, pedestrians, and potential impediments. To guarantee the dataset's representativeness and richness, a rigorous curation process involved picking a significant number of data points. This stage was crucial in providing the model with a wide range of railway scenarios, enabling a thorough learning process, and enhancing the model's ability to identify and classify things. To be more precise, a predefined number of samples, or 80% of the dataset for training and 20% of the dataset for testing and validation. Accuracy, precision, recall, F1-score, and recall were a few of the important metrics used to develop a comprehensive picture of the model's performance. Finally, a critical component of the statistical analysis of the object detection model was the careful distribution of 20% of the sample for testing and validation. This methodology yielded significant insights into the performance characteristics of the model and guided its iterative refinement and optimization, guaranteeing the model's efficacy and dependability within the complex and ever-changing context of railway crossings. The selected methodology, which is distinguished by its careful testing, validation, and dataset allocation processes, is crucial in obtaining significant insights into the complex performance features of the created object identification model. This methodology has revealed the model's advantages and disadvantages through a systematic evaluation procedure, offering a thorough grasp of how it functions in various scenarios found at railway crossings.

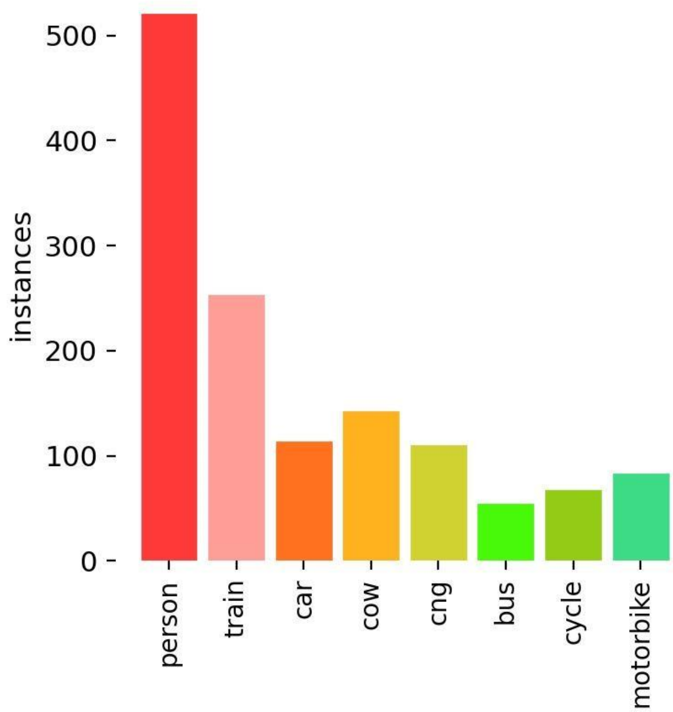


Figure 3.2.3: Instances of different classes

Table 3.3.1: Number of datasets

Class	**Images**
Person	651
Train	326
Cow	163
Car	157
CNG	123
Cycle	88
Motorbike	86
Bus	77

Almost 550 are raw data which is captured on railway crossings, open areas, and roads and almost 400 images are collected from online. Which is a large-scale dataset with different trainable classes, were a total of 949 and they comprise Person, Train, Cow, Car, CNG, Cycle, Motorbike, and Bus classes.

3.4 Proposed Methodology

In this research, the approach consists of 5 consecutive steps (Figure 3.4.1). First, I collected a dataset of images that I will use in the training and validation of the detection system. Then, I pre-processed this entire dataset by annotation. I trained the YOLO object detection models with the dataset, which I split for training. Using the dataset, I split for validation, I validated the real-time detection performance of trained models and evaluated the results.

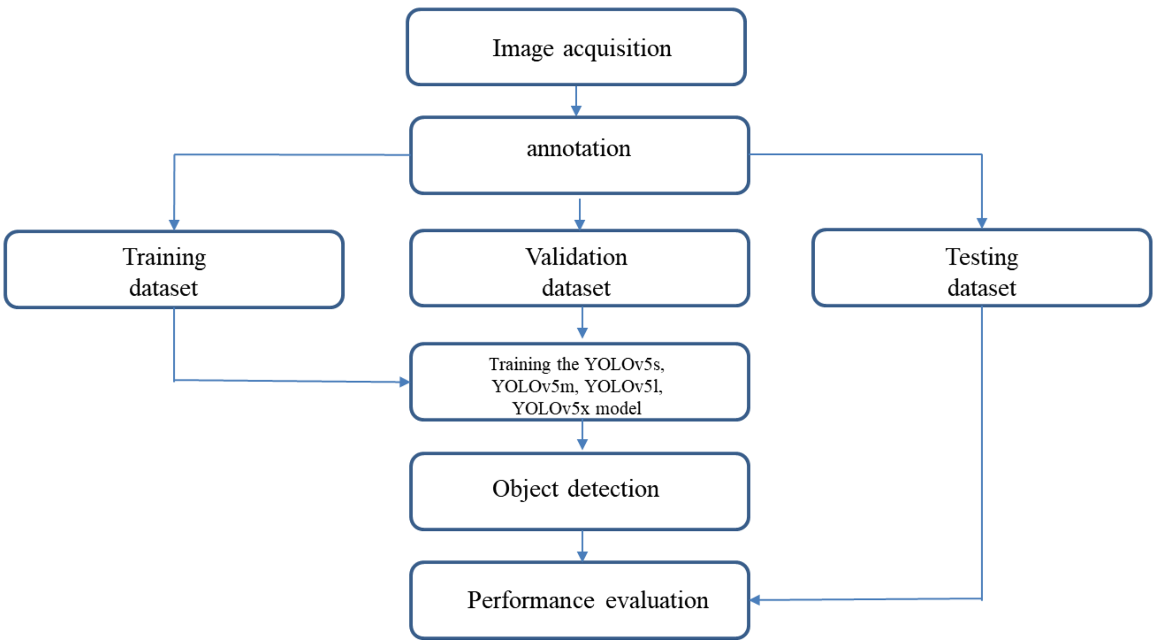


Figure 3.4.1: Flowchart of the research method

Data Preprocessing

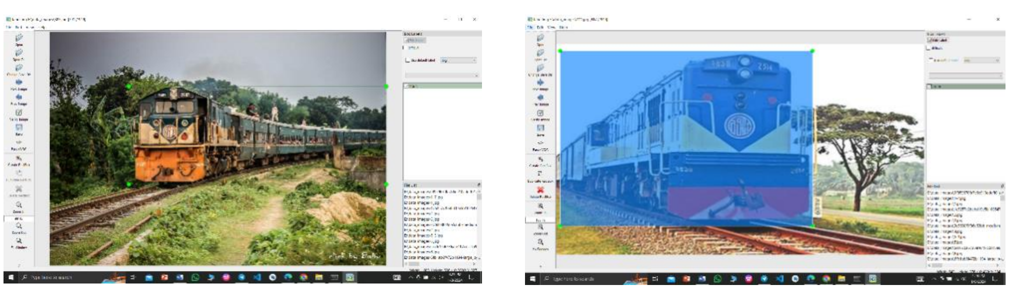


Figure 3.4.2: Image annotation software

I subjected those images, which I divided into different folders for training and validation, to data annotation in PASCAL VOC format.

The annotation process consists of marking the object to be detected in each image by taking it into a rectangle. Information about the objects marked on each image is stored in an XML file with the same name as the related image. Figure 3.4… shows the content of the sample annotation file of an image. For each image, I created an XML file with the same name as the image, which contains the annotated information.


```xml
<annotation>
  <folder>data_images</folder>
  <filename>411.jpg</filename>
  <path>E:\data_images\411.jpg</path>
  <source>
    <database>Unknown</database>
  </source>
  <size>
    <width>790</width>
    <height>452</height>
    <depth>3</depth>
  </size>
  <segmented>0</segmented>
  <object>
    <name>train</name>
    <pose>Unspecified</pose>
    <truncated>0</truncated>
    <difficult>0</difficult>
    <bndbox>
      <xmin>8</xmin>
      <ymin>14</ymin>
      <xmax>550</xmax>
      <ymax>390</ymax>
    </bndbox>
  </object>
</annotation>
```


Figure 3.4.3: Sample annotation XML file for an image

YOLOv5 uses a YAML file to instruct a parser on how to build a model.

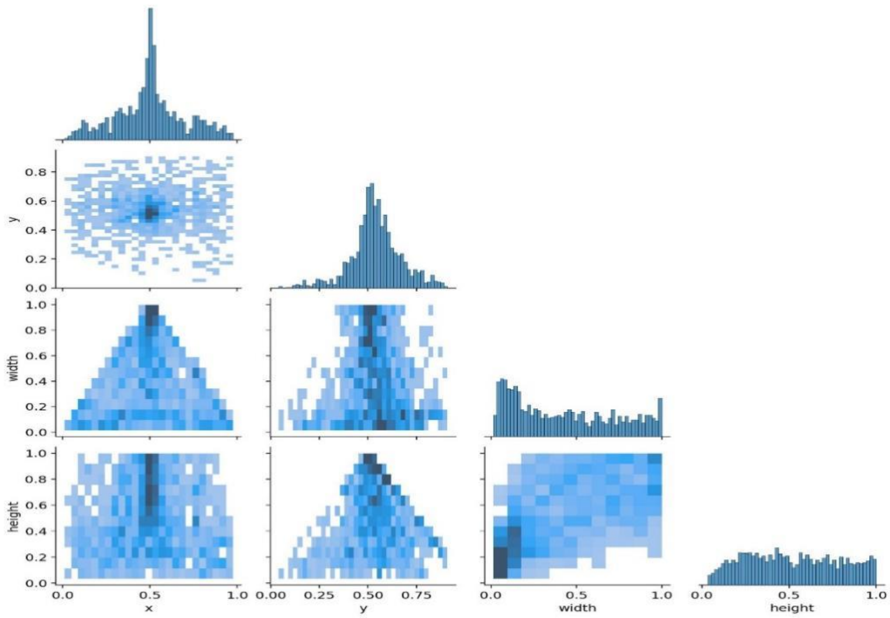


Figure 3.4.4: Relation between the position, width, and height of instances of the dataset

YOLO

With its creative approach to real-time detection, the YOLO (You Only Look Once) paradigm has changed the object detection landscape and is considered a fundamental advancement in the field. YOLO's primary innovation is its all-encompassing approach, which divides each image into a grid and processes it in a single pass. To produce a single, cohesive end-to-end solution, each grid cell simultaneously assumes the tasks of class probability and bounding box predictions. This divergence from conventional two-step procedures, which obviates the necessity for region proposals, leads to unmatched inference efficiency. The clever division of the image in YOLO's grid-based methodology is another example of its brilliance. This grid technique allows for a comprehensive loss function in addition to allowing the model to recognize objects of different sizes within the same architecture with efficiency. This function ensures that the model can detect objects precisely and accurately by considering the subtleties of bounding box position, size, and class probabilities. An important improvement is the addition of anchor boxes, which improves the model's capacity to forecast bounding box dimensions accurately at various object scales. As the most recent step in the YOLO series' evolutionary path, YOLOv5 introduces notable improvements in design, training methods, and overall performance.

YOLOv5, developed by Ultralytics, uses the CSPDarknet53 backbone architecture, which was purposefully selected to maximize feature extraction. This simplified architecture is a purposeful improvement, not just an evolution, designed to be efficient in training and real-world inference settings.

One unique aspect of YOLOv5 is its innovative approach to training methods. By introducing new techniques like label smoothing and auto-augmentation, the model improves its capacity for generalization and strengthens its resistance to changing data conditions. The versatility of YOLOv5 is demonstrated by the variety of model sizes it offers, ranging from YOLOv5s to YOLOv5x. With this range, users can choose the ideal balance between model size, computational resources, and accuracy, considering their unique resource limitations and application requirements. A further indication of YOLOv5's dedication to performance excellence is its intense focus on enhanced data augmentation during training. By exposing the model to a wide range of scenarios, methods such as mosaic augmentation and multi-scale training help it become more robust and adaptive in real-world situations.

The model's ease of use and technical strength are both demonstrated by its interoperability with PyTorch, a popular deep-learning framework. Because of its versatility, the model can be used with a wide range of hardware, meeting a variety of deployment requirements. To sum up, YOLOv5 is a powerful and adaptable solution for real-time object recognition across a wide range of applications and dynamic environments, and it is a tribute to the iterative development of the YOLO series.

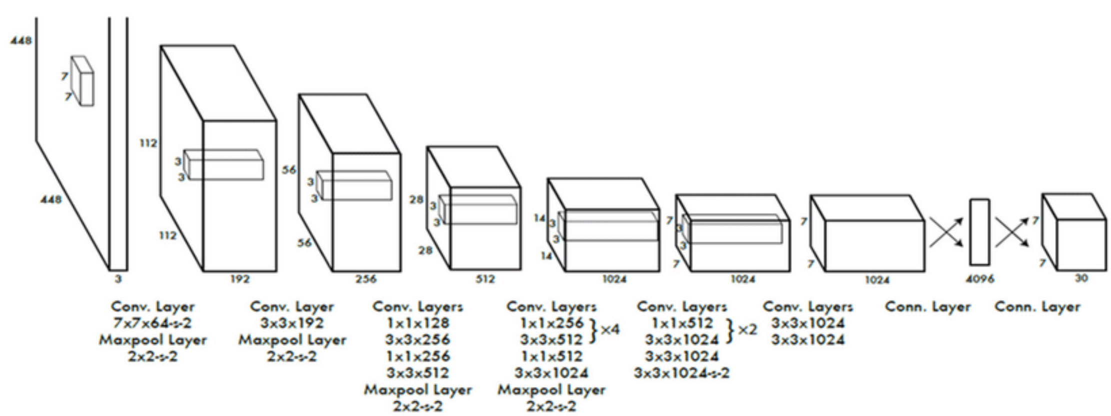


Figure 3.4.5: YOLO Model

YOLOv5

YOLOv5, a deep learning architecture that is well-known for its effectiveness in real-time object detection tasks, was the main tool used in this study's research technique for object recognition in railway crossings. The entire procedure was designed to make the most of YOLOv5's capabilities, expediting the creation of a reliable and accurate model for object identification and classification in the dynamic environment of railroad crossings.

The YOLOv5 model uses three components: CSPDarknet53 as the backbone, Path Aggregation Network (PANet) as the neck, and three YOLO heads. The data is initially input to the CSPDarknet component for feature extraction, which is then passed to the PANet component, which contains a feature pyramid network (FPN). The PANet component consists of multiple layers that mix and combine image features. The resulting feature representation is then passed to the three-head layer, which predicts the bounding boxes, confidence scores, and class probabilities for each detected object.

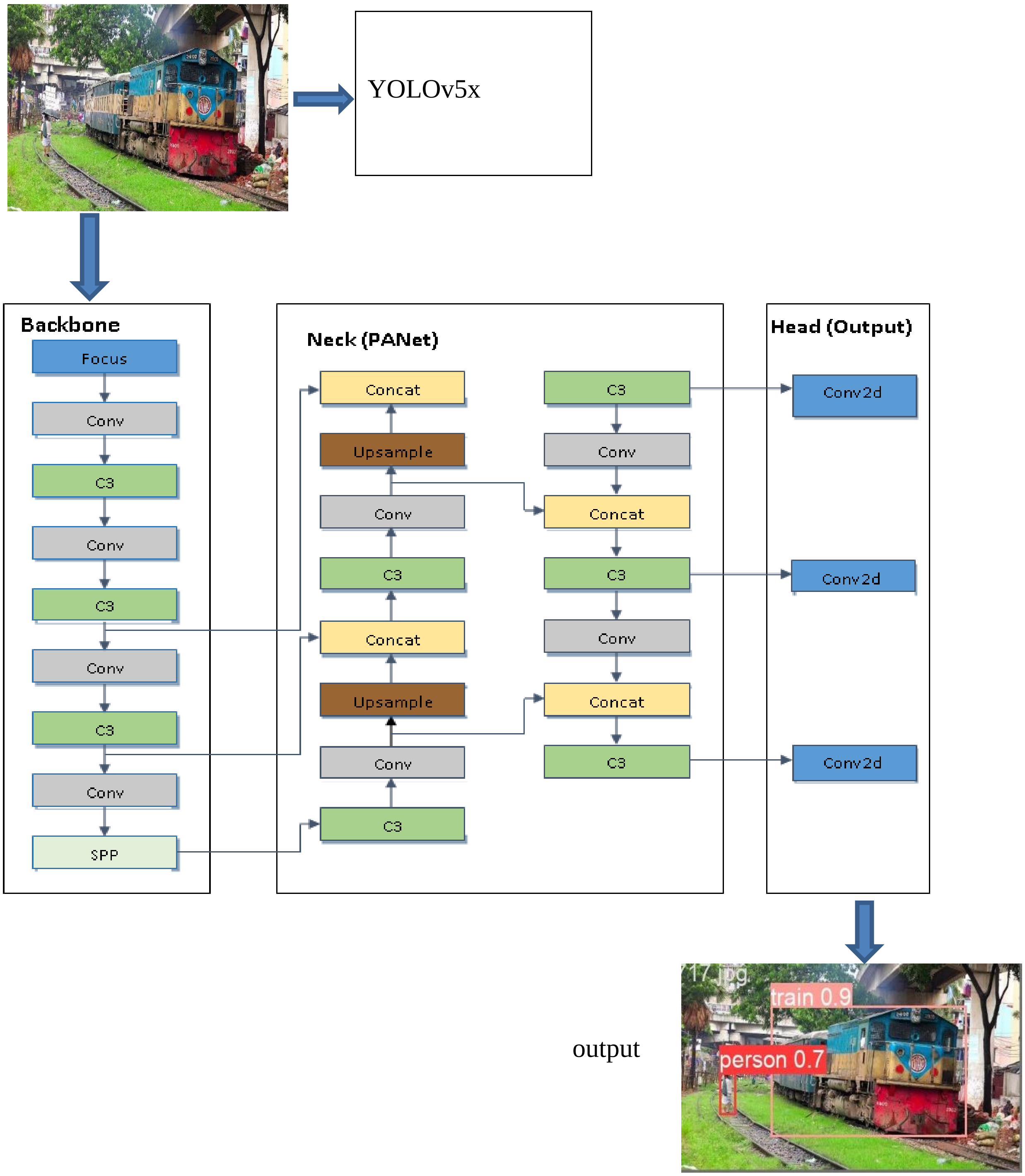


Figure 3.4.6: YOLOv5 Architecture of the proposed model

In this study, I used 4 versions of YOLOv5, which are named according to the model size and complexity, as small, medium, large, and extra-large.

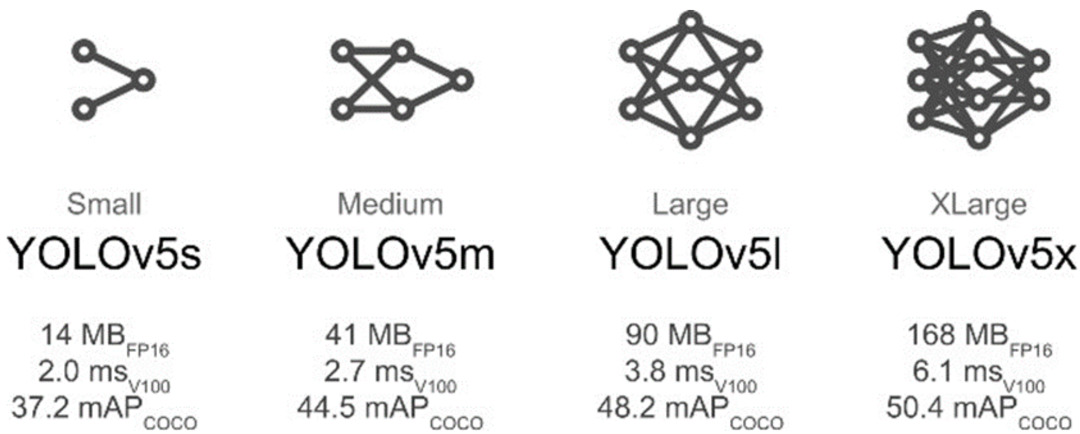


Figure 3.4.7: Different complexity of YOLOv5 version

3.5 Performance Evaluation

It is calculated as TP (True Positive) if IoU is greater than the defined threshold value, and as FP (False Positive) if it is small. Precision, recall, and mAP performance metrics were calculated using Eqn. 1,2 and 3 respectively with the obtained TP, TN (True Negative), FP, and FN (False Negative) values.

Precision: The ratio of correctly predicted positive instances to the total predicted positives. Precision equations shown in (Eqn.1)

$$\text{Precision} = \frac{TP}{TP+FP} \qquad \text{———} \qquad 1$$

Recall: The ratio of correctly predicted positive instances to the total actual positives. Recall equations shown in (Eqn.2)

$$\text{Recall} = \frac{TP}{TP+FN} \qquad \text{———} \qquad 2$$

The average precision (AP) is defined as the region under the precision–recall curve, which denotes the average accuracy.

$$\text{AP} = \int_0^1 P(R)dR$$

The mean average precision (mAP) is the aggregate of the AP for different categories and N is the number of categories. mAP equations are shown in (Eqn.3)

$$\text{mAP} = \frac{1}{N}\sum_{i=1}^{N} APi \qquad \text{Eqn.3}$$

N: Number of quires, **AP**: Average precision

3.6 Implementation Requirements

The system requirements are:

● Windows 10 or above OS

● SSD/HDD of minimum space of 100GB

● Minimum 8 GB RAM

Developing Tools requirements**:**

● Python with Tensor flow

● Google Colab

● Jupyter Notebook

● Browser

CHAPTER 4
EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Experimental Setup

Within the experimental framework designed for object recognition at railroad crossings, a thorough investigation takes place, covering the full procedure of utilizing trained models to identify and categorize things in this particular setting. The first stage is gathering a large dataset that includes various items and scenarios that are specific to railroad crossings. After gathering data, model preprocessing becomes the main focus and outlines several crucial activities. Installing necessary packages is part of this process, which also guarantees compatibility with the deep learning frameworks of choice—such as PyTorch or TensorFlow. The pre-trained models are precisely defined, and every image data is resized to fit the model criteria, further fine-tuning the system for peak performance. Crucially, the next step involves training the deep learning model by using the training and validation datasets to iteratively improve the model's parameters. The objective of this phase is to improve the model's capacity to precisely identify and categorize objects in the vicinity of the railroad crossing. Strict training procedures are used to guarantee the model's competence in practical situations. During the experimentation process, the testing data is subjected to a customized preprocessing procedure that aims to align it with the needs of the trained model, specifically about cross-entropy configuration. This strategic alignment guarantees that the model is ready to recognize and classify items efficiently throughout the testing phase. The capabilities of the model are then presented in the results section, which shows how accurate it is at classifying pertinent things in the railway crossing environment. This all-inclusive approach, which includes training, testing, preprocessing, and data gathering, creates a strong foundation for item detection at railroad crossings. The focus on each step enhances the model's overall effectiveness and promises accurate and dependable performance in real-world applications where accuracy and safety are crucial.

4.2 Experimental Results & Analysis

Train the model

For the first time experiment of the proposed system, I used YOLOv5s which is the smallest and fastest. Moreover, 758 images were for training, and 189 images were for validation.

For the second time, I used YOLOv5m. The experiment uses the same dataset and also 758 images were for training, and 189 images were for validation.

For the fourth time, I used YOLOv5l as the same procedure as those models.

For the third time, I used YOLOv5x. It is a large model of the YOLOv5 family. This experiment uses the same dataset.

Performance of YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5x

4.2.1 shows the results of precision, recall, mAP@50, and mAP@50-95 for all classes on the first dataset with model YOLOv5s. Table 4.2.1 shows the results for the same dataset with model YOLOv5m. Table 4.2.1 shows the results for the same dataset with model YOLOv5l. Table 4.2.1 shows the results for the same dataset with model YOLOv5x.

Table 4.2.1: Performance of YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5x

Class	YOLOv5s				YOLOv5m				YOLOv5l				YOLOv5x			
	Precision	Recall	mAP @0.5	mAP @0.5: 0.95	Precision	Recall	mAP @0.5	mAP @0.5: 0.95	Precision	Recall	mAP @0.5	mAP @0.5: 0.95	Precision	Recall	mAP @0.5	mAP @0.5: 0.95
Person	0.875	0.846	0.927	0.666	0.853	0.903	0.937	0.713	0.822	0.931	0.95	0.772	0.894	0.855	0.956	0.779
Train	0.824	0.863	0.904	0.631	0.768	0.858	0.906	0.667	0.794	0.863	0.913	0.684	0.88	0.756	0.909	0.7
Car	0975	0.784	0.949	0.664	0.912	0.932	0.947	0.722	0.959	0.944	0.962	0.776	0.97	0.892	0.956	0.799
Cow	0.804	0.795	0.917	0.664	0.772	0.864	0.926	0.671	0.756	0.886	0.942	0.743	0.869	0.818	0.943	0.756

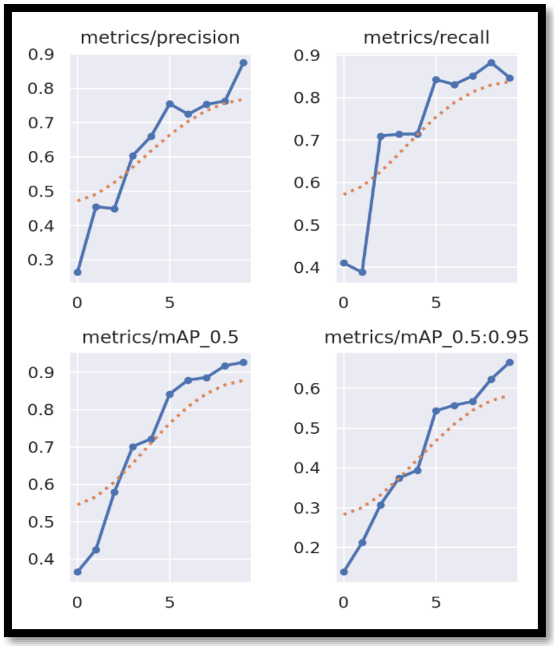
CNG	0.907	0.885	0.947	0.708	0.952	0.911	0.957	0.704	0.953	0.916	0.981	0.836	1	0.905	0.989	0.831
Bus	0.582	0.857	0.831	0.668	0.479	1	0.823	0.609	0.473	1	0.928	0.735	0.557	0.897	0.926	0.743
Cycle	0.936	0.875	0.93	0.702	1	0.755	0.956	0.752	0.953	0.839	0.972	0.806	0.933	0.584	0.937	0.783
Motorbike	0.973	1	0.995	0.653	0.944	0.955	0.989	0.724	0.91	1	0.982	0.818	1	0.99	0.995	0.876

Accuracy in Terms of mAP@50 and mAP@50-95

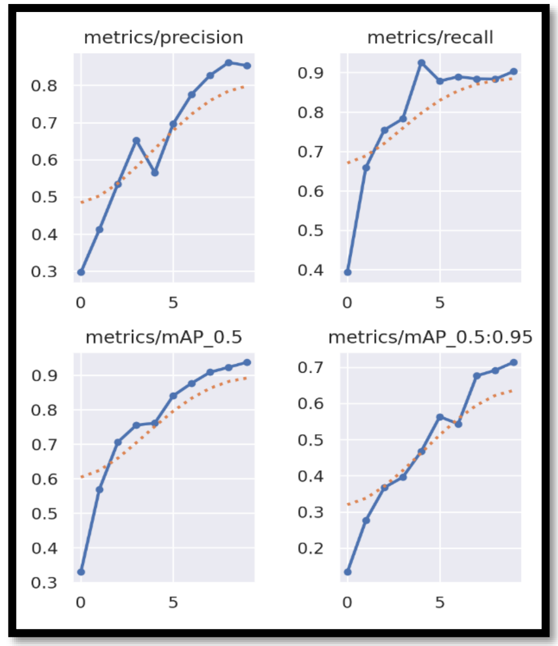
For mAP50 and mAP50-95 are evaluation metrics used in object detection tasks. mAP50 measures the mean average precision at an intersection over union (IoU) threshold of 0.5, while mAP50-95 measures the mean average precision across IoU thresholds ranging from 0.5 to 0.95. These metrics assess the accuracy of object detection models by calculating the precision and recall of predicted bounding boxes compared to the ground truth. The choice of metric depends on the task's specific requirements and the desired balance between precision and recall.

Comparison of models YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5x

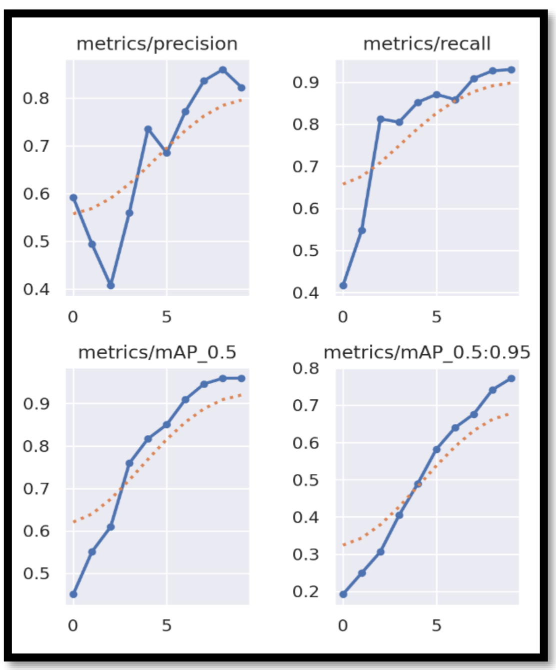
The four models were tested on the same datasets. For the dataset with 949 images, the difference between the model of precision, recall, mAP@50, and mAP@50-95 are shown in Figure 4.2.1 YOLOv5x is a bit more stable during the training.



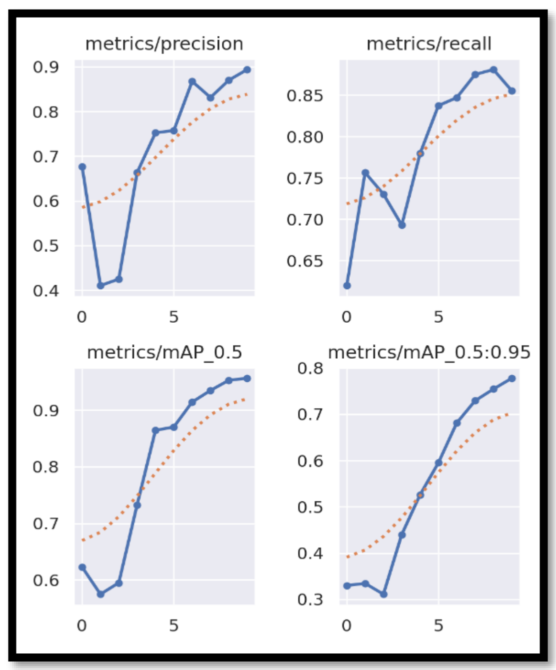
YOLOv5s



YOLOv5m



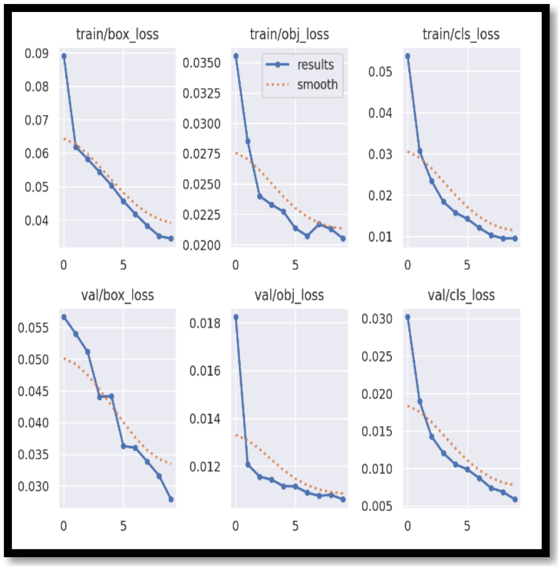
YOLOv5l

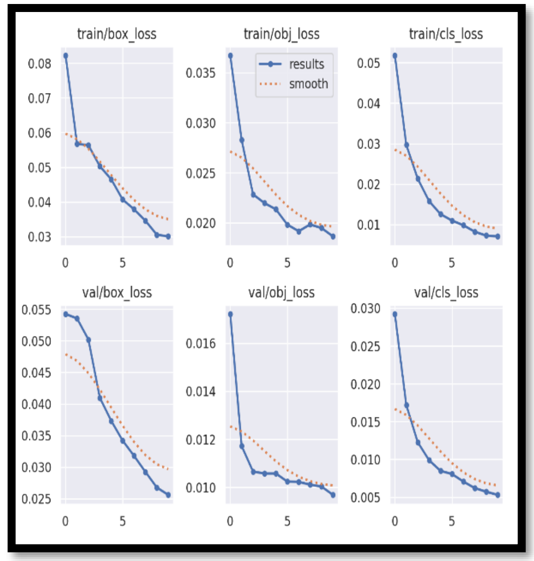


YOLOv5x

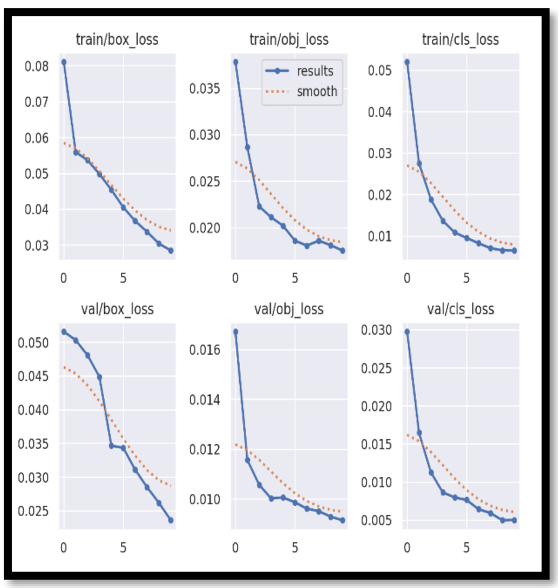
Figure 4.2.1: Comparison of performance metrics for YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x

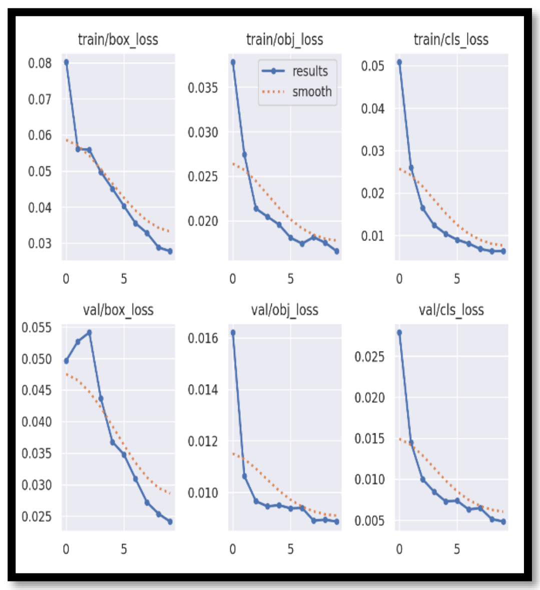
In the context of evaluating the YOLOv5 model for object detection in railway crossings, precision, recall, and loss are key metrics that provide insights into the model's performance and training dynamics. The number of affirmatively identified instances that are truly correct is called precision. An improvement in the YOLOv5 model's capacity to precisely identify and classify items in photos of railroad crossings is indicated by a rise in precision. The model becomes better at reducing false positives as precision increases, meaning that when it predicts an object, it is almost certainly right. Conversely, recall assesses the model's capacity to recognize every pertinent occurrence of an object in the dataset. A higher recall rate indicates that the model is getting better at identifying every instance of an object in photos taken at railroad crossings. A greater recall suggests a decrease in false negatives, proving the model's ability to correctly detect and categorize a greater percentage of real items in the scenarios. It shows the difference between the ground truth labels and the model's predictions. As the training process progresses, a declining loss signifies that the model is gradually approaching a more precise depiction of the fundamental patterns inherent in the data. A well-performing model shows a downward trend in loss and an upward trend in precision and recall. A decreasing loss indicates that the model is improving its predictions and gaining knowledge from the training set, while an increase in precision and recall points to increased sensitivity and accuracy, respectively.





YOLOv5s YOLOv5m



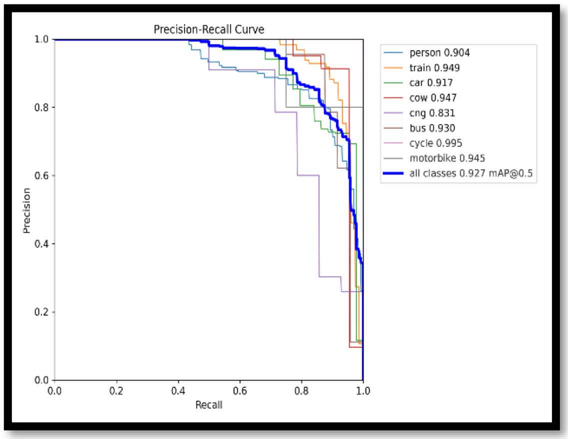


YOLOv5l YOLOv5x

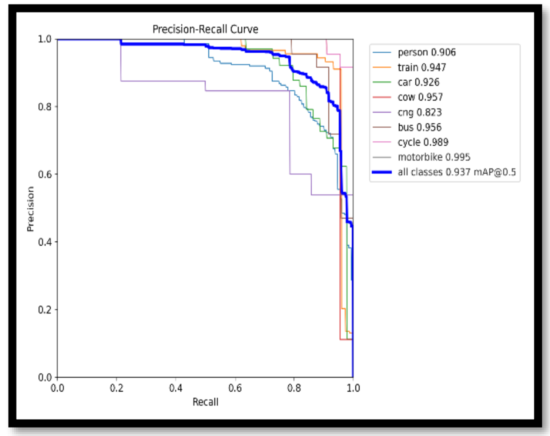
Figure 4.2.2: Model metrics of YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x

Figure 4.2.2 shows the loss value during the training. There are those values for the dataset with 949 images with the model YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. The chart shows that the YOLOv5x is very fast and stabilizes the learning process. Figure 4.2.

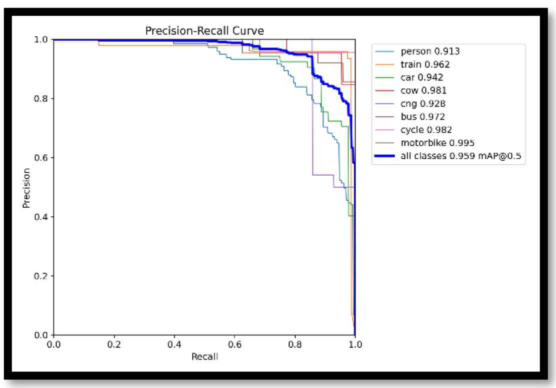
also shows the loss related to the predicted bounding box, an object, and the class loss for the four models YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x on the dataset with 949 images.



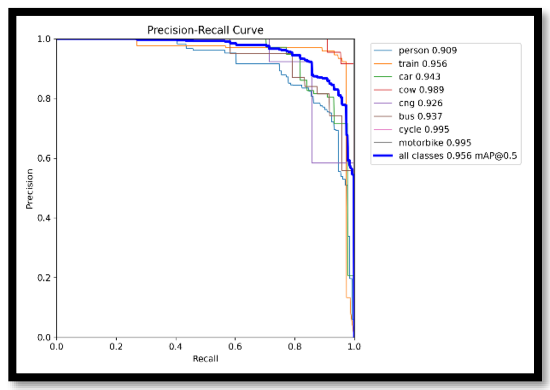
YOLOv5s



YOLOv5m



YOLOv5l



YOLOv5x

Figure 4.2.3: Precision and Recall of YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x

Figure 4.2.3 shows the precision and recall values of YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. Precision and recall are defined in Equations (Eqn.1) and (Eqn.2).
For precision compare the results of YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x from Table 4.2.1. Where YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x all classes had 87.5%, 85.3%, 82.2%, 89.4%. From the comparison, YOLOv5x has more true positives.

For recall comparing the results of YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x all classes had 84.6%, 90.3%, 93.1%, and 85.5%. From the comparison, YOLOv5m has more true positives.

A detailed examination of how changing the decision threshold affects the model's accuracy and recall values can be seen in the precision-recall trade-off graph. The graph displays a convincing trade-off situation with precision on the y-axis, which goes from 0 to 1, and recall on the x-axis, which goes from 1 to 0. Recall decreases and precision increases as the decision threshold rises, climbing from the bottom-left to the top-right corner. Initially, the model with a lower threshold finds more positive occurrences at the expense of precision, leading to higher recall. Nevertheless, recall suffers as precision increases as the threshold rises.

As we can see in Figure 4.2.4 the predicted label and the confidence of the prediction are shown on the image and most of the images have been predicted with high accuracy.

Original Pictures

Prediction results

Figure 4.2.4: Output of YOLOv5x

4.3 Discussion

A YOLOv5 model was proposed for real-time object detection. Table 4.3.1 compares the results of real-time object detection for all classes for YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x, on the same datasets—a total of 758 images for training and 189 images for validation. Based on experiment results, YOLOv5x achieved the highest average mAP.

Table 4.3.1: Tests with YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5x for same datasets.

Model	**Precision**	**Recall**	**mAP@50**	**mAP50-95**
YOLOv5s	0.875	0.846	0.927	0.666
YOLOv5m	0.853	0.903	0.937	0.713
YOLOv5l	0.822	0.931	0.95	0.772
YOLOv5x	0.894	0.855	0.956	0.779

Table 4.3.2 shows the performance of the level of complexity of a variety of models including layers, parameters, and GLOPS. In this experiment, YOLOv5x contains the most layers, parameters, and GLOPS which are 322, 86220517, and 203.9 followed by the YOLOv5s, YOLOv5m, YOLOv5l.

Table 4.3.2: Comparison of the Layers, Parameters, and GLOPS

Model	**Layers**	**Parameters**	**GLOPS**
YOLOv5s	157	7031701	15.8
YOLOv5m	212	20881221	47.9
YOLOv5l	267	46145973	107.8
YOLOv5x	322	86220517	203.9

The thorough assessment of the object detection model provides significant insights into its behavior across a range of decision thresholds and confidence levels, especially when done through a variety of graphical analyses like recall-confidence, precision-recall trade-off, precision-confidence curve, and F1 score-confidence trade-off. The model's sensitivity is shown as a function of confidence threshold variation in the recall-confidence graph. The trade-off shown by the observed relationship between strong recall and lower confidence criteria and vice versa highlights how flexible the model is to various application requirements. The model exhibits a fine balance between comprehensiveness and precision as it gets more conservative and prioritizes higher confidence at the expense of some recall. The inverse link between precision and recall is shown in the precision-recall trade-off graph. The model captures all positive cases (recall) at the cost of minimizing false positives (precision). This trade-off is critical in situations where recall and precision have different relative relevance, directing the model's behavior according to particular application needs. The precision's reliance on confidence levels is explored via the precision-confidence curve. Precision rises as the model's confidence threshold does, suggesting a lower chance of false positives. This detailed knowledge is essential for adjusting the behavior of the model to match the intended degree of prediction certainty. Last but not least, the F1 score-confidence trade-off graph offers a comprehensive assessment of the model's overall effectiveness. To achieve a high F1 score, the ascending phase shows an initial balance between recall and precision. Nevertheless, the F1 score falls as the model gets stricter and emphasizes greater confidence, underscoring the delicate balance needed for the best possible model performance.

CHAPTER 5

IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

5.1 Impact on Society

With its innovative approach to object detection at railroad crossings, this project has the potential to significantly alter society by tackling important safety issues in the transportation industry. The project is centered on applying deep learning approaches to sophisticated object recognition techniques. Its goal is to develop a reliable system that can identify objects accurately and quickly, hence making the training environment safer and more secure. The fundamental goal of the initiative is to greatly improve safety precautions at railroad crossings. The suggested object recognition system seeks to lower the danger of collisions between trains and cars or people, avoid accidents, and save lives by utilizing state-of-the-art deep learning techniques. Identifying things at railway crossings promptly can be crucial in preventing possible mishaps and fostering a feeling of safety for both commuters and nearby communities. The effort possesses the capability to enhance overall efficiency and optimize transportation procedures, in addition to safety concerns. At railroad crossings, a well-installed object detection system can promote more efficient traffic movement, cut down on delays, and lower the risk of accidents. Enhancing the overall functionality of public areas and transportation infrastructure, commuters and pedestrians stand to gain from a more secure and efficient transit network. The consequences of the project go beyond object detection and computer vision. The program advances computer vision innovation and continues the journey of technological advancement by creating and executing sophisticated models for railway safety. For comparable systems in allied fields like traffic management, surveillance, and critical infrastructure security, the study findings and techniques used could provide a basis. The program is in line with the general objective of promoting accessibility and inclusivity in public areas from a societal standpoint. Advanced object recognition technology makes sure that people with different abilities can use transportation systems safely and with accessibility. This inclusivity makes transportation more accessible, user-friendly, and equitable, especially for people who struggle to navigate public spaces or have visual impairments. To sum up, the influence of this program extends beyond the narrow field of railroad safety and permeates all public areas and transit systems. Through the resolution of important safety issues, enhancement of transportation effectiveness, and promotion of inclusivity, the project has the potential to transform safety protocols, advance technology integration, and make a substantial contribution to the development of a society that is safer and more accessible to everyone. The results of this project could have a significant impact on how we think about accessibility and safety in public areas, which would eventually benefit society as a whole.

5.2 Impact on Environment

This project has significant, though indirect, environmental benefits. It is largely motivated by the need to improve communication accessibility for those with hearing impairments. The project's concentration on utilizing digital platforms and cutting-edge technologies highlights its unintentional positive impact on environmental sustainability, even if the main purpose is to facilitate communication. The development of an advanced object detection system for railroad crossings is in line with both safety requirements and a larger commitment to environmentally friendly behaviors by utilizing effective digital communication techniques. The project's advocacy for the use of digital platforms as a communication tool demonstrates its deep impact. The project promotes digital communication techniques by developing an advanced object-detecting system, which greatly reduces the requirement for in-person contact and physical travel. This shift to digital platforms actively lowers carbon emissions related to travel and commuting while simultaneously improving communication accessibility. The project catalyzes reducing the environmental footprints associated with physical attendance, such as fuel use and road congestion, in industries like employment and education where remote solutions are supported.

The project's emphasis on digital interpretation and communication is a critical aspect of its environmental impact. Through the project's promotion of digital platforms and applications for object detection at railway crossings, the necessity for conventional paper-based communication approaches is successfully reduced. In addition to streamlining communication, the move to digital techniques directly addresses environmental issues including trash production and deforestation, which is a big step towards sustainable communication practices. The project's advances in computer vision algorithms and deep learning techniques have wider ramifications than their immediate uses. Energy-efficient computer solutions are unintentionally promoted by the search for more efficient hardware and software. The project's commitment to developing effective object detection systems at railroad crossings indirectly lowers the overall energy consumption and carbon footprint associated with the computationally intensive processes needed for these kinds of applications. More environmental sustainability is sparked by the project's active promotion of inclusive practices and increased public awareness. It emphasizes how environmental stewardship, societal inclusion, and technical developments are interrelated. The project's inadvertent environmental contributions demonstrate its diverse effects, establishing it as a key player in guiding society toward a more sustainable and inclusive future. The initiative affects society in addition to technology breakthroughs. The project builds on increased environmental consciousness by promoting inclusivity and awareness. A more varied and inclusive society is better equipped to promote and enact eco-friendly policies and practices in a range of industries. The project is in line with environmentally friendly methods and serves as an example of how technological improvements may be crucial in advancing sustainability and inclusivity on a larger scale. It demonstrates how technology innovation may have a positive impact on environmental sustainability and societal inclusion by fusing traditional communication methods with digital ones. Thus, by raising public knowledge and encouraging participation, the project catalyzes greater environmental sustainability. Although communication accessibility is the project's main goal, its unintentional yet considerable environmental contributions highlight its overall influence. More environmental sustainability is sparked by the project's active promotion of inclusive practices and increased public awareness. The project is in line with ecologically friendly practices by advocating for digital communication methods and influencing social viewpoints. It also demonstrates how technological improvements can be crucial in promoting inclusivity and sustainability on a larger scale.

5.3 Ethical Aspects

The development of an object identification system for railway crossings through the integration of deep learning techniques demands careful consideration of ethical concerns at every stage of the project. This technology-driven program raises important ethical issues that require serious consideration in the pursuit of improving railway safety. To ensure that the object detection system is developed responsibly and inclusively, ethical principles must be strictly followed throughout the entire process, from data collection to system deployment. The first section of the ethical framework focuses on participant permission and data privacy. Explicit approval from pertinent parties is essential in the case of railway crossings, where safety is of utmost importance. Respectful data acquisition and use are guided by privacy laws and ethical principles. Taking precautions to secure and protect personal information is necessary, especially when dealing with potentially susceptible groups. Securing participant identities throughout data collection, storage, and analysis requires strong data anonymization protocols. Stakeholder participation and active inclusivity are required by an ethical approach. The project should include key players who are knowledgeable about railway crossings, pertinent stakeholders, and communities impacted by railway infrastructure. It becomes morally required to obtain informed consent and to actively participate while being mindful of cultural quirks. Involving the community in the planning, design, and assessment stages improves the system's usability, efficacy, and general acceptability in addition to being a procedural necessity. Given the diversity of railway crossings, the object detection system's ethical design needs to take various signaling techniques, cultural quirks, and regional variances into account. In order to guarantee that the system correctly recognizes a broad range of objects and to promote equitable representation and inclusion for all stakeholders involved in railway safety, ethical issues go beyond technological functioning. The presence of algorithmic bias in deep learning applications necessitates careful consideration. To reduce biases, ethical practices require that training datasets be carefully chosen and balanced. Ethical requirements include diversity, representation, and the eradication of bias. To find and address any potential biases, the system's performance must be regularly evaluated across a range of demographic groupings. The object detecting system's accessibility and usefulness are also considered from an ethical standpoint. Prioritizing flexibility and simplicity will help the user interface accommodate users of different skill levels. For people with a variety of needs to utilize the system effectively and independently, it must be interoperable with assistive devices. Deep learning models must operate transparently in order to promote accountability and confidence. It is important for users and stakeholders to comprehend how the system understands and identifies objects at railroad crossings. The development of methods that clarify the model's decision-making process and improve understanding ought to be the focus of ethical endeavors. Proactively evaluating how the object detection system will affect society is a crucial component of an ethical strategy. Important ethical considerations include making sure that use is responsible, preventing harmful use, and maintaining accountability. Ethical duty includes keeping an eye on the system's implementation and acting quickly to reduce any unintentional harmful consequences on society. Researchers make sure that the creation and implementation of the object detection system for railroad crossings conform to the values of fairness, confidentiality, inclusiveness, and respect for the rights of all parties involved by thoroughly addressing these ethical issues. This moral approach protects user welfare and encourages moral behavior, which adds to the project's overall beneficial effects on society. By managing the nexus of technology and ethics, this methodology guarantees conscientious innovation for improved rail safety.

5.4 Sustainability Plan

The railway crossing object detection project is purposefully developed to embrace sustainability on multiple fronts, considering aspects related to the environment, economy, and society. This time, the focus is only on the railway safety context, which promotes a thorough and long-lasting effect. The project, which is at the forefront of environmental sustainability, uses digital communication platforms instead of traditional paper-based approaches. This change reduces the amount of paper used, as well as the amount of waste produced and the carbon emissions caused by physical materials and transportation. The idea is in line with environmentally sustainable practices by incorporating sophisticated object identification algorithms into railway safety procedures. The project's improvement of safety protocols at railroad crossings makes a major contribution to social inclusion. In railway environments, the real-time object detection technology promotes smooth communication and teamwork, guaranteeing equitable involvement and interaction for everybody. Communities become more resilient and peaceful as a result of this inclusivity, which also supports equality, diversity celebration, and social harmony. The project's concentration on technology development specifically related to railway safety is a fundamental component of its sustainability. The application of state-of-the-art object identification algorithms improves the accuracy and efficacy of safety protocols at railroad crossings. In addition to addressing today's safety concerns, these technological developments foster continued innovation in the field of railroad infrastructure and safety. The project's scalability and flexibility to various railway settings strengthen its sustainability. Extensive datasets encompassing diverse railway crossings, signaling techniques, and geographical variances are used for training, which makes the system naturally culturally and contextually adaptive. This flexibility guarantees the project's ongoing applicability and efficacy in a variety of railway situations. The project's sustainability is based in large part on its economic feasibility. Through the implementation of a real-time object detection system, the project offers affordable ways to improve railway safety without depending on expensive alternatives. This guarantees accessibility. In addition to guaranteeing the project's longevity, its economic viability promotes broad adoption for the improvement of railroad safety. Collaborative efforts with multiple stakeholders, such as railway specialists, technology developers, policymakers, and local communities, are essential to the project's viability. These partnerships enable the sharing of expertise, the use of best practices, and a dedication to continuous upkeep and enhancement of the object detection system. Including a range of stakeholders guarantees a comprehensive strategy for guaranteeing safety at railroad crossings. Maintaining sustainability is a dynamic process that requires a steadfast dedication to continuous improvement in railway safety protocols. Frequent input is essential for identifying areas for improvement, resolving new difficulties, and making sure the object detection system is up to date and functional, especially from train operators and local populations. This iterative process ensures that the project will remain relevant and long-lasting in the constantly changing field of railway safety. In conclusion, the object detection project for railroad crossings is complex and sustainable, considering social, economic, and environmental concerns. By utilizing digital communication, improving technological capabilities, guaranteeing scalability and adaptability, promoting economic viability, encouraging collaboration, and consistently enhancing safety measures at railway crossings, the initiative seeks to create a safer and more inclusive society.

CHAPTER 6

SUMMARY, CONCLUSION, RECOMMENDATION AND IMPLICATION FOR FUTURE RESEARCH

6.1 Summary of the Study

By using the YOLOv5 model to object detection in railway environments, safety and efficiency at railway crossings have been significantly improved. The YOLO (You Only Look Once) model is well-known for its real-time object identification speed and accuracy, which makes it a great option for railway applications. The YOLOv5 model brings state-of-the-art object identification technology to the field of railroad safety. Its capacity to evaluate the entire image at once and quickly and accurately identify several objects improves the overall safety protocols at railroad crossings. Version 5, which is based on the YOLO architecture, improves the model's efficiency and makes it more appropriate for complex railway applications. The effectiveness of the model in identifying and categorizing objects helps to quickly and precisely evaluate possible hazards at railroad crossings. YOLOv5's ability to digest information in real-time is one of its main advantages. In fast-moving railroad environments, the model's speed guarantees that objects—such as cars, pedestrians, and obstructions—are identified and categorized with exceptional accuracy, enabling prompt action. The YOLOv5 model exhibits flexibility in various railway scenarios. The model is capable of adapting to diverse environments with ease, having been trained on datasets that cover a range of railway crossings, signaling techniques, and geographical variances. Its flexibility to adapt to various railway settings adds to its effectiveness in guaranteeing safety. One important component of the YOLOv5 model's longevity is its economic feasibility. The model encourages accessibility without depending on pricey alternatives by offering a practical means of improving railway safety through real-time object detection. The model's long-term effects are enhanced by its widespread adoption, which is encouraged by its economic viability. Collaborative efforts with a variety of stakeholders, such as railway experts, technology developers, policymakers, and local communities, are essential to ensuring the model's sustainability. Through knowledge sharing, best practices, and continuous maintenance, these partnerships provide an all-encompassing strategy for railway safety. Maintaining sustainability requires a dedication to continuous development. Frequent input is especially helpful in identifying areas for improvement, addressing new issues, and making sure the YOLOv5 model is still relevant and useful in meeting changing railway safety regulations. This is especially true for feedback from local communities and train operators. In conclusion, the application of the YOLOv5 model for object detection at railroad crossings is a cutting-edge approach. Its real-time accuracy, flexibility, economic feasibility, cooperative nature, and dedication to ongoing development make it an invaluable tool for promoting safety protocols at railroad crossings. The YOLOv5 model is a cutting-edge technological advancement that makes the railway environment safer and more effective.

6.2 Conclusions

In conclusion, applying YOLOv5 to railway safety is a ground-breaking development. The model is a vital tool for handling complicated issues at railroad crossings because of its real-time accuracy, adaptability, and economic feasibility. The YOLO design allows for YOLOv5's enhanced speed, which improves danger recognition by enabling rapid and accurate object detection in dynamic railway situations. Its long-term impact on safety is influenced by its adaptability to various railway situations, which is fueled by training on many datasets to assure efficiency across scenarios.

Finally, we have concluded that after applying different pre-trained models of the YOLOv5 series on real-time data, the YOLOv5x model generates the highest average of precision at 89.4%, recall at 85.5%, and mAP at 95.6%. So YOLOv5x is the most stable method followed by YOLOv5s, YOLOv5m, and YOLOv5l. The accuracy of the object detection system improved by increasing the dataset size with new images added and data augmentation. In this way, it detects objects accurately at railway crossings and reduces accident rates.

6.3 Implication for Future Study

Given the project's future direction, it is strategically necessary to add more sophisticated capabilities to the existing framework, most notably Explainable AI and the Internet of

Things (IoT). The aforementioned changes are intended to enhance not only the system's functions but also make a significant contribution to the domains of automation and interpretability—two critical areas in the rapidly changing field of railway safety technology. One of the most important steps toward automating the railway crossing object detection system is the incorporation of IoT. The integration of IoT devices into the infrastructure confers onto the system the ability to gather data in real-time from its environment. This data infusion improves the system's responsiveness and flexibility by allowing it to react independently to sudden changes in the environment. At railway crossings, automated decision-making procedures that are based on real-time data streams from IoT devices might result in quicker and more efficient reactions to possible threats. Furthermore, it becomes clear that including Explainable AI will be a crucial component of the work to come. It becomes increasingly important to comprehend the machine learning models' decision-making processes as the system develops and gets more complex. Explainable AI approaches are useful because they offer clear insights into the models' prediction-making process. This interpretability is a tool for stakeholder trust-building as well as helping to understand future projections. Explainable AI's capacity to provide context for the system's activities encourages transparency, which is essential for stakeholders to feel comfortable interacting with and depending on the system. This project will essentially take a two-pronged approach in the future: it will use Explainable AI to improve interpretability and IoT to automate tasks. It is envisaged that these cutting-edge capabilities would catapult the railway crossing object identification system into a more complex and open domain. The system's ability to adjust to changing difficulties is ensured by the combination of Explainable AI with IoT, which builds stakeholder confidence as technology progresses.

References

[1] Li, K.;Wan, G.; Cheng, G.; Meng, L.; Han, J. Object Detection in Optical Remote Sensing Images: A Survey and A New Benchmark 2019. ISPRS J. Photogramm. Remote Sens. 2020, 159, 296–307. [CrossRef]

[2] Heitz, G.; Koller, D. Learning Spatial Context: Using Stuff to Find Things. In Proceedings of the Computer Vision—ECCV 2008; Forsyth, D., Torr, P., Zisserman, A., Eds.; Springer: Berlin, Heidelberg, 2008; pp. 30–43.

[3] Learning Rotation-Invariant Convolutional Neural Networks for Object Detection in VHR Optical Remote Sensing Images | IEEE Journals & Magazine | IEEE Xplore. Available online: https://ieeexplore.ieee.org/document/7560644 (accessed on 31 December 2023).

[4] Razakarivony, S.; Jurie, F. Vehicle Detection in Aerial Imagery: A Small Target Detection Benchmark. J. Vis. Commun. Image Represent. 2016, 34, 187–203. [CrossRef]

[5] Zhu, H.; Chen, X.; Dai, W.; Fu, K.; Ye, Q.; Jiao, J. Orientation Robust Object Detection in Aerial Images Using Deep Convolutional Neural Network. In Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, Canada, 27–30 September 2015; pp. 3735–3739.

[6] DLR—Earth Observation Center—DLR Multi-Class Vehicle Detection and Orientation in Aerial Imagery (DLR-MVDA). Available online: https://www.dlr.de/eoc/en/desktopdefault.aspx/tabid-12760/22294_read-52777 (accessed on 31 December 2023).

[7] Liu, Z.; Yuan, L.; Weng, L.; Yiping, Y. A High Resolution Optical Satellite Image Dataset for Ship Recognition and Some New Baselines. In Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods, Porto, Portugal, 24–26 February 2017. [CrossRef]

[8] Xia, G.-S.; Bai, X.; Ding, J.; Zhu, Z.; Belongie, S.; Luo, J.; Datcu, M.; Pelillo, M.; Zhang, L. DOTA: A Large-Scale Dataset for Object Detection in Aerial Images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018.

[9] Zou, Z.; Shi, Z. Random Access Memories: A New Paradigm for Target Detection in High Resolution Aerial Remote Sensing Images. IEEE Trans. Image Process. 2018, 27, 1100–1111. [CrossRef]

[10] ITCVD Dataset—University of Twente Research Information. Available online: https://research.utwente.nl/en/datasets/itcvddataset accessed on 31 December 2023).

[11] Scottish Index of Multiple Deprivation 2020. Available online: https://www.gov.scot/collections/scottish-index-of-multipledeprivation-2020/ (accessed on 31 December 2023).

[12] Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Wu, E. Squeeze-and-Excitation Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019.

[13] Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. arXiv 2020, arXiv:1910.03151. Remote Sens. 2023, 15, 614 16 of 17

[14] Hou, Q.; Zhou, D.; Feng, J. Coordinate Attention for Efficient Mobile Network Design. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021.

[15] Chen, S.; Tan, X.; Wang, B.; Hu, X. Reverse Attention for Salient Object Detection. In Computer Vision—ECCV 2018; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Springer International Publishing: Cham, Switzerland, 2018; Volume 11213, pp. 236–252. ISBN 978-3-030-01239-7.

[16] Lin, H.; Cheng, X.;Wu, X.; Yang, F.; Shen, D.;Wang, Z.; Song, Q.; Yuan,W. CAT: Cross Attention in Vision Transformer 2021. In Proceedings of the 2022 IEEE International Conference on Multimedia and Expo, Taipei City, Taiwan, 11–15 July 2022.

[17] Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.

[18] Kasper-Eulaers, M., Hahn, N., Berger, S., Sebulonsen, T., Myrland, Ø. & Kummervold, P. E. (2021). Detecting heavy goods vehicles in rest areas in winter conditions using YOLOv5. Algorithms, 14(4), 114.

[19] Malta, A., Mendes, M., & Farinha, T. (2021). Augmented reality maintenance assistant using yolov5. Applied Sciences, 11(11), 4758.
[20] Wan, J., Chen, B., & Yu, Y. (2021). Polyp Detection from Colorectum Images by Using Attentive YOLOv5. Diagnostics, 11(12), 2264.
[21] Yao, J., Qi, J., Zhang, J., Shao, H., Yang, J., & Li, X. (2021). A real-time detection algorithm for Kiwifruit defects based on YOLOv5. Electronics, 10(14), 1711.
[22] Yao, J., Qi, J., Zhang, J., Shao, H., Yang, J., & Li, X. (2021). A real-time detection algorithm for Kiwifruit defects based on YOLOv5. Electronics, 10(14), 1711.
[23] Córdova, M., Pinto, A., Hellevik, C. C., Alaliyat, S. A. A., Hameed, I. A., Pedrini, H., & Torres, R. D. S. (2022). Litter detection with deep learning: A comparative study. Sensors, 22(2), 548.
[24] Jiang, X., Hu, H., Qin, Y., Hu, Y., & Ding, R. (2022). A real-time rural domestic garbage detection algorithm with an improved YOLOv5s network model. Scientific Reports, 12(1), 16802.
[25] Lin, F., Hou, T., Jin, Q., & You, A. (2021). Improved YOLO based detection algorithm for floating debris in waterway. Entropy, 23(9), 1111.
[26] Zang, H., Wang, Y., Ru, L., Zhou, M., Chen, D., Zhao, Q., ... & Zheng, G. (2022). Detection method of wheat spike improved YOLOv5s based on the attention mechanism. Frontiers in Plant Science, 13, 993244.
[27] Liu, G., Mao, S., Jin, H., & Kim, J. H. (2019, February). A robust mature tomato detection in greenhouse scenes using machine learning and color analysis. In Proceedings of the 2019 11th International Conference on Machine Learning and Computing (pp. 17-21).
[28] Qiang, L., Jianrong, C., Bin, L., Lie, D., & Yajing, Z. (2014). Identification of fruit and branch in natural scenes for citrus harvesting robot using machine vision and support vector machine. International Journal of Agricultural and Biological Engineering, 7(2), 115-121.
[29] Rahnemoonfar, M., & Sheppard, C. (2017). Deep count: fruit counting based on deep simulated learning. Sensors, 17(4), 905.

Plagiarism Report

Defense1

ORIGINALITY REPORT

25%	21%	15%	11%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	publisher.unimas.my Internet Source	3%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	3%
3	Submitted to Huntington Beach Union High School District Student Paper	2%
4	www.mdpi.com Internet Source	2%
5	www.researchgate.net Internet Source	2%
6	Submitted to Daffodil International University Student Paper	1%
7	Dahang Wan, Rongsheng Lu, Sailei Wang, Siyuan Shen, Ting Xu, Xianli Lang. "YOLO-HR: Improved YOLOv5 for Object Detection in High-Resolution Optical Remote Sensing Images", Remote Sensing, 2023 Publication	1%